



Protocol: Post-basic Technical and Vocational Education and Training (TVET) Interventions to Improve Employability and Employment of TVET Graduates in Low- and Middle-income Countries

Janice Tripney, Mark Newman, Kimberly Hovish, and Chris Brown

Submitted to the Coordinating Group of:

<input type="checkbox"/>	Crime and Justice
<input checked="" type="checkbox"/>	Education
<input type="checkbox"/>	<input type="checkbox"/> Disability
<input checked="" type="checkbox"/>	International Development
<input type="checkbox"/>	<input type="checkbox"/> Nutrition
<input type="checkbox"/>	Social Welfare
<input type="checkbox"/>	Other:

Plans to co-register:

<input checked="" type="checkbox"/>	No		
<input type="checkbox"/>	Yes	<input type="checkbox"/> Cochrane	<input type="checkbox"/> Other
<input type="checkbox"/>	Maybe		

Date Submitted: 10 December 2011

Date Final Revision Submitted: 19 June 2012

Approval Date: 27 June 2012

Publication Date: 1 September 2012

BACKGROUND

The Problem, Condition or Issue

In the twenty-first 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 has been uneven, however. Whilst some developing countries, particularly China and India, have considerably improved their standing in the global economy, many non-globalising developing countries have fared much worse; many 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 means there are more skilled workers in the world than available prospects. Simultaneously, global unemployment is on the increase. Recent shocks provoked by the international financial crisis exposed severe weaknesses within the global economic system which rapidly spread to the employment sector, triggering a global jobs crisis. The global economy has substantially reduced its capacity to add new jobs.

Youth have been particularly hard hit by the economic crisis. The youth unemployment rate rose sharply during this period, from 11.6 to 12.7 per cent, and, in the absence of economic growth and development, this pattern is likely to continue (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 (UN, 2012; UN/DESA, 2011). Furthermore, worldwide, many young people are underemployed (in part-time, temporary or low productivity jobs), and/or work in low quality, poorly paid jobs. The vast majority of the world's youth work in the informal economy, many under poor working conditions or in hazardous forms of work. In many countries, young women are much more likely to be unemployed 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 also experiencing a 'youth bulge' (that is, have two-thirds of their populations under the age of 30), a demographic shift which compounds what are already severely limited opportunities for integrating youth into the labour market. One billion young people will reach employment age within the next decade (ILO, 2012).

Rising levels of youth unemployment and underemployment impose heavy costs on individuals and their families, society and the economy. Both prolonged absence from the labour market and poor quality/low productivity 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 opportunities for economic growth. Increasing numbers of youth are moving to urban areas in search of employment (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 and Palmer, 2010). The political and policy communities in many low- and middle-income countries 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 (World Bank, 2007). The latest UNESCO Global Monitoring Report highlights the policy importance now being placed on higher-order skills and the central place they play in the global knowledge-based economy, with regard to poverty reduction, economic growth and social stability (UNESCO, 2010a). This shift in priorities is exemplified by the theme of the 2012 Education for All (EFA) Global Monitoring Report (due to be published in September 2012), which will focus on TVET by exploring the links between skills programmes and employment (UNESCO, 2010b). 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 programmes. For instance: in the Philippines around \$200 million was invested by the government in TVET in 2002 (Péano et al, 2008); in Indonesia about \$80 million dollars was invested by the Asian Development Bank and \$35 million by the government in 2008 (ADB, 2008); in Uganda, participation in TVET programmes has increased by almost 50% within a decade (IMF, 2010).

The Intervention

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 and 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). 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). A great diversity of TVET models can be found worldwide, with no internationally accepted set of definitions of the different types that can be distinguished. For the purposes of this systematic review, the following definitions have been used:

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; delivered at (usually) 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.

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/Employment.

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; also called dual-training programmes.

The organisation of TVET varies widely, both between and within countries. Brief overviews of three relevant TVET programmes in low- and middle-income countries follow.

Jóvenes en Acción (Youth in Action) was a Colombian training programme funded by the national government, and financed with loans from the World Bank and the Inter-American Development Bank. The programme was aimed at socio-economically disadvantaged young people in seven cities across the country with the aim of helping people who were hardest hit by the recession of 1998. It reached 80,000 young people (approximately 50 percent of the target population) and was given to various cohorts over a period of four years (2002-2005). Participants underwent a total of 6 months of training, formed of 3 months classroom-based training (various manual and administrative courses were offered) and 3 months unpaid on-the-job training. Training took place at private and non-profit training centres and legally registered companies. Private training institutions had to participate in a bidding process to be able to participate in the programme. *Source: Attanasio et al. (2008, 2009, 2011)*

The ongoing Technical and Vocational Vouchers Program (TVVP), which began in Kenya in 2008, offers young people the opportunity to undertake vocational and technical training through awarded vouchers which completely, or almost completely, cover the costs of training courses. Participants can use the vouchers to undertake training in either public or private institutions, on courses such as motor-vehicle mechanics and hairdressing. (NB: The impacts of participating in training are examined.) *Source: Hicks et al. (2011)*

Retraining programmes for laid-off workers (all ages) implemented following new national policy measures in China have been evaluated. The training was provided primarily by city labour bureaus, but also by private sector institutions (including city and district employment training centres, colleges, universities, and secondary technical schools). The quality of programmes varied across training institutions. Duration of training averaged one to three months. Most trainees contributed to the cost of their training. Courses on offer (some of which accommodated labour market demand for certain skills) included computing, cooking, beauty, massage and hair cutting, sewing, toy making, management, repairs and driving. *Source: Bidani et al. (2002, 2005, 2009)*

How the Intervention Might Work

This logic model provides a simple representation of the relationships among (i) the resources that are invested; (ii) the activities that take place; and (iii) the benefits or changes that result, as a sequence of events.

Inputs	Activities	Outputs	Intermediary outcomes	Outcomes/impacts
money, staff, equipment, etc.	TVET programmes (classroom training, on-the-job internships, etc.)	completion of a TVET programme	intermediary outcomes represented by the general construct <i>employment</i> : for example, job searches, job applications, job interviews intermediary outcomes represented by the general construct <i>employability</i> : 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	outcomes represented by the general construct <i>employment</i> : for example, gaining initial employment; maintaining employment (including making transitions between jobs and roles within the same organisation to meet new job requirements); obtaining new employment (e.g., through promotion); self-employment (starting a new business or expanding one); working hours; payments (i.e., earnings, wages, salary or income)

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 policy makers 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-sufficiently within the labour market (McQuaid & Lindsay, 2005). Particularly outside policy circles, 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.

For all interventions, whether job creation is additional or not is a key issue. Most projects will have both positive and negative impacts. All of these should be taken into account in order to assess the net difference that results from the TVET intervention, over and above what would have taken place anyway – i.e., estimates of gross employment outcomes should be adjusted to take into account displacement and substitution effects.

- 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.

Whilst recognising that determining the ‘additionality’ of any employment effects is methodologically very challenging, it is important to be aware that even if effective in terms of employment outcomes, a TVET programme might not generate any additional employment if substitution or displacement effects are present. In both instances, TVET might only be affecting who gets employed, not the level of employment.

Why it is Important to do the Review

Despite considerable international evidence, the body of literature taking stock of the evaluation evidence on TVET and young people is relatively small. 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. A minority extend previous reviews by explicitly considering the impacts for young people and/or for populations in developing and transition countries. A discussion of some of these reviews follows.

Kluge and Schmidt (2002) compared the results of a sample of European impact evaluations of ALMPs implemented between 1983 and 1999, and compared their results to the U.S. programmes previously studied by Heckman et al. (1999). Their analysis suggests mixed programme effects across categories of intervention and target population. Young workers were found to be the most difficult group to assist among the unemployed. Kluge (2006) followed this up with a meta-analysis of European ALMPs in the later 1990s and 2000s. More recently, Card et al. (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, programmes for 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 et al. (2003) who synthesised findings from 15 publicly-funded training programmes in the United States to measure programme 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 programmes for youth, although again these findings are mostly from advanced countries.

There have been a number of reviews based on programmes in the World Bank’s ‘Youth Employment Inventory’ database (a global inventory of more than 400 projects to support

young workers in over 90 countries) including those focused specifically on young people and/or developing countries (see, for example, Betcherman et al., 2004, 2007; Fares and Puerto, 2009; Katz, 2008; Puerto, 2007; Stavreska, 2009). Betcherman et al. (2007) summarised information on a large number of international programmes supporting young people in their early years in the labour market. Although the largest concentration of included interventions is from OECD countries, there are also substantial numbers of programmes introduced in the largely middle-income countries of Eastern Europe and Central Asia, Latin America and the Caribbean. They found that training is the dominant form of intervention used to help young people improve their employment situation. For training as a whole, mixed results are reported, although their analysis shows an increasing incidence of positive impacts from programmes that offer multiple services, i.e., combinations of vocational training, job search assistance, entrepreneurial services, and so forth. Training programmes for youth seem to have a more positive impact in developing countries than in developed countries. More recently, Angel-Urdinola et al. (2010) analysed the main design features of non-publicly provided ALMPs in Arab-Mediterranean Countries, with a specific focus on initiatives targeted at youth. Benchmarked against international best practices, assessment of the programmes covered in the inventory reveals that the majority lack the necessary mix of design features that make such programmes effective.

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 clear grounds for concentrating this review solely on the effects of TVET programmes on youth in low- and middle-income countries. There is also motivation for this systematic review from a methodological perspective. Many of the existing reviews, although synthesising evidence from programmes which appear relevant to our review, are not based on a systematic search and several use a 'vote-counting' approach to synthesis. These are problems which this review will aim to remedy; in so doing, adding value to the existing body of research on this topic.

OBJECTIVES

The overall aim of this systematic review is to help policy-makers, practitioners and academics understand the available evidence on technical and vocational education and training (TVET) interventions in order to guide programmatic efforts to increase employment and employability amongst youth in low- and middle-income countries (LMICs). The review will answer the following questions:

- What are the effects of different models of post-basic technical and vocational education and training (TVET) interventions on the employment and employability outcomes of graduates, 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, a main objective of the review is to systematically gather and synthesise the relevant evidence, showing variation in treatment effects, magnitude of effects, and the relationship between magnitude and mode of TVET. In addition, evidence of differential effects will be explored: for trainees with different characteristics (i.e., in relation to gender, education level, work experience, length of current employment status), by programme characteristics (i.e., in relation to type of TVET, duration of training, programme location, sector/industry, and voluntariness) and by labour market conditions. Possible reasons for varying or conflicting results will be discussed. To enhance the usability of findings, a second objective is to examine, to the extent possible, the applicability and transferability of the interventions from the study setting to the local setting. A final objective is to identify gaps in the literature and suggest potential avenues for future research.

METHODOLOGY

This review will be conducted in accordance with current Campbell Collaboration Guidelines on Systematic Review Methods.

Criteria for Including Studies in the Review

Studies must meet the following eligibility criteria to be included in the systematic review.

Types of study designs

To be included, a study must use an experimental or quasi-experimental design. Eligible designs include those in which the authors use a control or comparison group and in which: participants are randomly assigned (using a process of random allocation, such as a random number generation); a quasi-random method of assignment has been used and pre-treatment equivalence information is available regarding the nature of the group differences (and groups generated are essentially equivalent); participants are 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 are non-randomly assigned, but statistical methods have been 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 these studies may include youth receiving no treatment, treatment as usual, or an alternative treatment. No restriction will be placed on duration of follow up.

Types of participants

Countries vary considerably in their definition of youth. The standard United Nations definition of youth as those belonging to the 15-24 years age group will apply to this review (United Nations, 1992).

Participants will have the following characteristics:

- *Age*: Young people aged 15 - 24 years: either the whole sample (because intervention is targeted at youth) or average age lies between 15 and 24 years (in cases where the intervention is open to a wider age range);
- *Geographical location*: From low- or middle-income countries (as defined by the World Bank: see Appendix 1);
- *Gender*: Male and/or female (i.e., both dual- and single-sex studies are eligible for inclusion in the review);
- *Target group*:
 - i. Any employment status at time of service receipt (i.e., not in paid employment or in paid full- or part-time employment);
 - ii. Any skills level, prior experiences, achievements or level of qualification.

Types of interventions

Inclusion in the proposed systematic review is restricted to TVET interventions with the following characteristics:

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

Types of outcome measures

To be included, a study must assess intervention effects on at least one eligible outcome variable. Qualifying outcome variables are those that fall in the following general construct categories: (a) employment and (b) employability.

- primary outcomes represented by the general construct *employment*: for example, gaining initial employment; within-organisation mobility (moving between roles within the same organisation); maintaining employment (including making transitions between jobs and roles within the same organisation to meet new job requirements); obtaining new employment (employment transitions between organisations); promotion; self-employment (starting a new business or expanding one); working hours; and payment levels (i.e., earnings, wages, salary or income);
- intermediary outcomes represented by the general construct *employment*: for example, job searches, job applications, job interviews;
- intermediary 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) are eligible for inclusion. However, despite the importance of this issue, we anticipate that the methodological difficulties in designing interventions that allow exploration of whether any employment effects are additional or not will result in the identification of few, if any, studies measuring effects on general (net) employment level.

Date, language and form of publication

UNESCO launched a long-term TVET programme from 2000, following the Second International Congress on Technical and Vocational Education, held in Seoul in 1999. For this review, the date of publication or reporting of the study must be 2000 or later. Eligible studies can be published in any language as long as they meet all other eligibility criteria. We will not exclude specific forms of publication, such as theses and dissertations.

Exclusion criteria

Outside the scope of the review is continuing professional development (CPD) for professionals (i.e., programmes designed to upgrade knowledge and skills of practitioners in the medical and other professions). Some evaluations of active labour market programmes (ALMPs) are eligible for inclusion in the review. To be excluded are those ALMPs where the

focus of the evaluation is not on vocational education or training, such as programmes providing assistance with job searching or those providing financial subsidies (an exception here would be the provision of financial assistance to attend/purchase training where trainees' participation in such training was then evaluated). Comprehensive multiple-service interventions (for example, combining on-the-job training with wage subsidies) are eligible for inclusion in the review. Excluded are evaluations of programmes promoting self-employment by providing technical assistance. To be eligible, interventions must not be targeted specifically at youth with particular special needs, such as learning disabilities, physical disabilities, emotional problems, or behavioural problems. Studies will be excluded if they are focused exclusively on vocational rehabilitation training programmes. These are considered a distinct sub-category of TVET-type initiatives, with separate, distinct systems in place, and with their own body of literature. It is reasonable not to include such initiatives with mainstream TVET. Studies in which youth with special needs have participated in mainstream TVET/skills training will be included, providing other criteria have been met. Lastly, studies using only perception measures (i.e., the views of employers and the workforce about their employability) are not eligible for inclusion in the review.

Search Strategy

A comprehensive search strategy will be used to search the international research literature for qualifying studies. Different types of sources will be searched, including sources with a particular focus on low- and middle-income countries (some of which were sourced from the Cochrane EPOC Group's list of sources relevant to LMICs:

<http://epocoslo.cochrane.org/lmic-databases>). A number of European-focused sources have been 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 1 for the World Bank list of low- and middle-income economies, grouped by region). The use of a wide range of sources is intended to capture both academic and 'grey' literature and reduce the omission of relevant studies, to ensure that our search is as unbiased as possible.

Documenting the search and selection processes

Review management software (EPPI-Reviewer 4) will be used to manage the entire review process (Thomas et al., 2011). Potentially relevant items identified through the electronic database search will be exported to EPPI-Reviewer and then each item manually screened for eligibility, with EPPI-Reviewer used to keep track of decisions made about each citation. For relevant studies identified through hand searching, it will be necessary to check whether the item is already in EPPI-Reviewer before proceeding to manually enter details for such items into EPPI-Reviewer.

All information retrieval and selection activities in the review will be documented and described in sufficient detail in the final report so that the processes can be replicated by other researchers (with summary flowcharts used to convey information, where relevant).

Based on the Cochrane PRISMA checklist for reporting results of searching and screening, the following information will be recorded: databases, database platforms, search strategy for at least one database, dates of search, time-frame, name of reviewer, and time spent on searching.

Electronic Search Strategy (electronic searching of databases)

A wide range of general and specialist electronic bibliographic databases will be searched. If during our search we come across others deemed relevant to the review, we will search those also. Electronic bibliographic databases to be searched include:

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/
- Scientific and Technical Egyptian Bibliographic Database (STEB) www.sti.sci.eg/enstinetdatabases.htm
- 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/

Search terms: A tailored search strategy will be 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 will combine a comprehensive list of search terms related to the intervention, outcomes and research design. Synonyms and wildcards will be applied as appropriate. Database thesauri will be consulted to ensure that all appropriate synonyms have been included. There will be no country or language restrictions to the search. A publication year filter to identify studies published since 2000 will be used. A draft search strategy for ERIC is presented in Appendix 2.

Hand Search Activities

To identify further studies for inclusion, the electronic search of databases will be supplemented by handsearching activities. We will conduct hand searches of websites/gateways, scan reference lists of included studies and relevant reviews, conduct forward citation tracking, contact experts and make requests through relevant networks.

Websites/Gateways:

Where there is a search facility that allows search terms to be entered, this will be used; otherwise, relevant sections (for example, those headed 'publications') will be 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.blids.ids.ac.uk/
- Eldis www.eldis.org/
- European Training Foundation www.etf.europa.eu
- 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

Reference lists:

The bibliographic information contained within the reference lists of included studies and relevant prior reviews will be scanned for studies that meet the eligibility criteria. The following reviews (deemed relevant as they review programmes that meet our selection criteria) will be searched (Angel-Urdinola et al., 2010; Betcherman et al., 2004, 2007; Bouillon and Tejerina, 2006; Card et al., 2010; Fares and Puerto, 2009; Godfrey, 2003; Ibarraran and Rosas Shady, 2008; Knowles and Behrman, 2005; Palmer et al., 2007; Spevacek, 2009; Van Adams, 2007). If during our search we come across others deemed relevant to the review, we will search those also.

Citation checking exercises:

Studies that have cited the included studies since their publication will be checked for relevance. All the hits from each citation search will be screened. Citation tracking will be performed through the ISI Web of Knowledge and Google Scholar.

Personal contacts:

We will email 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 will be made. (Authors and funding sources will also be contacted regarding the availability of translated versions of included studies.)

Networks:

Requests for relevant literature will be made through the following networks:

- Norrag: www.norrag.org
- UNESCO-UNEVOC e-forum <http://www.unevoc.unesco.org/forum.php>

Again, a specific request for assistance with the location of studies published in languages other than English will be made.

Search engines:

A keyword search using Google Scholar <http://scholar.google.co.uk/> will be undertaken. This has the potential to identify grey literature and studies prior to the results appearing in databases.

We will not undertake hand searching of individual journals or search for conference proceedings or dissertations separately.

Study inclusion decision making

Selection of primary studies will be based on the pre-developed selection criteria described above. The screening of literature for eligibility will be undertaken in two phases. For the majority of the source types (see Table 1), an initial round of screening based on titles and abstracts will be carried out, followed by an examination of the full-text of the study to determine eligibility. However, the full reports of items suggested by personal contacts or through requests made to members of relevant networks will be obtained and eligibility based on a reading of the full-text.

TABLE 1: STUDY INCLUSION DECISION MAKING

Source of records	Title and abstract screening	Full-text screening
Databases, websites, reference lists, forward citation tracking, search engines	√	√
Personal contacts, networks		√

Phase 1: Title & abstract screening: the first phase involves manual examination of the titles and (where available) abstracts of identified records. The relevance of each item will be

assessed by an individual reviewer who will work through the exclusion criteria hierarchically. The reviewers have three options (see Table 2).

TABLE 2: TITLE AND ABSTRACT SCREENING

Option	Action	
Option a: "exclude"	do not promote to next round of screening	When the flow process is stopped because the item meets a particular exclusion criterion, then the appropriate exclusion code will be recorded in EPPI-Reviewer.
Option b: "unsure"	promote to second round of screening	When the flow process is stopped because there is insufficient information within a title & abstract, either to exclude the item on a particular criterion or to proceed to the next criterion, the item will be marked as 'unsure'. It is likely that few studies will be excluded on criteria 4, 5 or 6, as information about outcomes, age of participants and study design is often not reported in abstracts. Items marked as 'unsure' will then be considered by both reviewers during the next phase of the screening process.
Option c: "provisionally include"	promote to second round of screening	Studies which, on the basis of their title and abstract, appear to meet the criteria for inclusion will be marked as a 'provisional include' and the full text retrieved.

In cases where the title and/or abstract are not in the English-language, the translation service offered by Google (<http://translate.google.com/>) will be used to translate the information into English and then screening against the selection criteria would proceed as normal.

In cases where only the title of the study is available (e.g., when scanning reference lists) reference within the wording of the title to (a TVET intervention) AND (a relevant employment-related outcome OR a term suggesting the study is an evaluation) will automatically warrant a full length review of the article.

The selection criteria will be piloted by two researchers who will screen a sample of reports independently and compare their results. Discrepancies will be resolved by further review of the respective titles and abstracts. This process will be repeated until a high level of consistency in application of the selection criteria is achieved. The remainder of the phase one screening will then be carried out by individual reviewers (i.e., single screening).

Before proceeding to the next phase of screening, the full length reports of all items marked 'unsure' or 'potentially relevant' will be retrieved. If these are not available online it, an inter-library loan request will be made.

Phase 2: Full-text screening: During this phase, detailed manual examination of full length papers will be undertaken to determine eligibility. Two reviewers will independently evaluate each study promoted from the first level of screening. The reviewers will then compare and

discuss their assessments. Any disagreements between the reviewers' decisions will be resolved by identification of the source of the disagreement, re-reading of the text and discussion. If a final decision cannot be reached, a third reviewer will be asked to reconcile differences.

Description of Methods Used in Primary Research

Studies to include in the review will employ experimental or quasi-experimental research designs that compare outcomes for an intervention group to those for a control or comparison group. The following two studies exemplify the methods likely to meet the eligibility criteria for the review.

Attanasio et al. (2008, 2009, 2011) used a randomised design in which the availability of training was randomly assigned among those who chose to apply for training (using a centralised information system). Randomisation took place within institutions. The data set included treatment and control groups for each of the seven cities covered by the programme. Baseline and one follow-up survey were completed. The authors report that there was close to full compliance with random assignment and programme effects were first estimated using simple comparison of weighted means. To increase the precision of the estimates, they then included controls for observable characteristics (pre-treatment characteristics and course fixed effects) to help control for remaining baseline differences.

The study conducted by Bidani et al. (2002, 2004, 2005, 2009) was based on a quasi-experimental design whereby participants and comparison groups were selected after the programme had commenced. The study used a survey designed and implemented by the Chinese Institute of Labor Studies and the World Bank. Respondents to the survey were selected from lists of laid-off workers who had received training (treatment group) and laid-off workers who had not received the training (control group) from lists provided by the Labor Bureaus from two large cities, as well as local training institutions. One post-test measurement was taken. Adjustments were made in the estimation process to account for the differences in the characteristics of the participant and comparison groups. Impact estimates were computed by several different estimators: ordinary least squares (OLS)/linear probability model; probit model; propensity score matching (nearest neighbour; five nearest neighbours; kernel; local linear regression).

Criteria for Determination of Independent Findings

Sometimes, several different reports (i.e., publications) about a single study exist. In such cases, the most relevant report should be identified (for example, the publication containing the most complete data set) and used as the main record. The coding, however, should be done from the full set of relevant reports (to be identified before beginning to code). There are also occasions where a single report describes more than one study. In these cases, each study should be coded separately [as if they had come from separate report]. Efforts will be made to identify all affiliations between studies/reports before coding commences.

Information on study sample sizes, intervention details, grant numbers, and so on, will be used to identify multiple reports from single studies, and multiple studies in single reports. The authors of the reports will be contacted if it is unclear whether reports and studies provide independent findings.

We have specified two general construct categories: employment and employability. We plan to extract effect sizes for all relevant outcomes and follow-ups measured within the same study. These will be separated according to (i) the general constructs they represent (employment, employability) and (ii) the key outcome areas represented by these general constructs. It is not felt to be appropriate to calculate an overall effect size for employment and employability. Rather, we anticipate conducting separate meta-analyses of the key outcome areas (i.e., change in employment status, income level, qualifications, job performance and so on). We expect to encounter statistical dependence among study outcomes in these key areas. For example, some studies may report two measures of job mobility or have measured changes in income level in different ways. Dependence may also occur when primary research studies present comparisons of multiple treatment groups to a common control group (when only a single outcome is observed).

In cases where a single study provides more than one effect size for a particular outcome construct, to ensure that each study only contributes one data point to the analysis for each outcome (i.e., to ensure independency of the findings) we will adopt the approach of either dropping or combining outcomes. Dropping outcomes: If there are only a few instances of a single study measuring an outcome construct in more than one way, we will select the construct that is most similar to those used by other studies in that category and retain only that particular effect size in the analysis. Combining outcomes: For construct categories where single studies measuring an outcome construct in more than one way is relatively common, all the effect sizes will be retained in the analysis and we will use the technique developed by Hedges et al. (2010) to estimate robust standard errors that account for the statistical dependencies.

In cases where multiple measurements were assessed in a single study over time, we aim to analyse the outcomes by the duration of each of the follow-up periods (e.g., outcomes at six months, two years, etc.). If a sufficient number of such studies are available, we will also analyse outcomes by investigating the change in effect size over time. However, in the event that synthesising effect sizes separately at different points of duration is not feasible (e.g., not all studies may use common follow-up durations) we will form reasonable ranges of follow-up duration (e.g., short term 1-3 months, long term 9-12 months, etc.) rather than discrete follow-up duration time points.

Details of Study Coding Categories

Two reviewers will independently evaluate each study to capture both substantive and methodological characteristics. A draft version of sections A-E of the coding tool (data extraction form) is included in Appendix 3. Eligible studies will be coded on features such as

year of publication, details of any linked publications, and so forth (section B), variables related to the characteristics of the study samples (section C), the nature of the intervention and its implementation (section D), and the study methods (section E). The coding process will also incorporate an assessment of the methodological quality and relevance of each study (section F). The different dimensions likely to be considered here include: study design (i.e., was the allocation mechanism or method of analysis able to control for selection bias), confounding (i.e., was the method of assignment and/or method of analysis executed adequate to ensure comparability of groups throughout the study and prevent confounding), contamination (spillovers and crossovers), performance bias, detection bias, construct validity, analysis and outcome reporting biases, statistical conclusion bias and external validity. We will draw upon the GRADE guidelines for grading the quality of evidence and strength of recommendations for practice and policy (Balshem et al., 2011). The statistical findings will be extracted (and effect sizes calculated) in section G of the tool.

We will collect data, if there is any, on net employment. Although net employment effects (i.e., those that take into account displacement and substitution effects) are seldom likely to be reported as an outcome, this issue has implication for interpreting programme effectiveness. Therefore, for each study we will record whether this issue has been addressed in any way, and/or whether the training took place in saturated/unsaturated markets, and so forth.

To enhance the usability of findings, we plan to draw upon an innovative approach to assessing applicability and transferability developed by Wang et al. (2005): where ‘applicability’ is defined as the extent to which an intervention process could be implemented in another setting, and transferability as the extent to which the measured effectiveness of an applicable intervention could be achieved in another setting. A range of intervention process and contextual factors will be collected accordingly.

We have used information in the primary study reports identified so far to identify potential effect size moderators; these include gender, education level, work experience, length of current employment status, type of TVET, duration of training, programme location (urban/rural), sector/industry, programme voluntariness and labour market conditions. We will assess the extent to which these moderating factors may be associated with heterogeneity in measured effects between studies.

The coding tool is based on previous EPPI-Centre tools, but has been modified where appropriate in accordance with Campbell guidelines. It also draws on previous work by Wilson et al. (2010). Reviewers will enter data directly into the EPPI-Reviewer 4 database, with the help of the coding manual where necessary. Piloting of the coding tool will be undertaken by members of the review team who will work independently on a random sample of eligible studies before meeting to compare their decisions. Reviewers will be retrained on any coding items that show discrepancies during this process and the coding manual adapted accordingly. This process will be repeated until a very high level of

consistency in reviewers' application of the codes is achieved (at which point the tool will be finalised). The remaining studies will be double-coded. Different combinations of two reviewers (KH/CB, JT/CB, JT/KH) will independently extract information from each study report and then come together to compare their decisions (all sections of the tool, except H). Any uncertainties and discrepancies will be resolved by discussion, further review of the respective study reports and, where necessary, consultations with a third reviewer (JT or MN). Section H, which relates to the statistical findings of the primary studies and the calculation of effect sizes, will be completed by two reviewers (JT and MN): again, each study will be double-coded. Where additional guidance on statistical issues is required, we will draw on the expertise of Alison O'Mara-Eves and James Thomas at the EPPI-Centre.

The reviewers will attempt to contact the authors of studies that are missing key data that is essential for the review. Where relevant studies have been published in languages other than English, authors and funding sources will be contacted regarding the availability of translated versions. Should these be unavailable, we will seek to identify (through Campbell) additional reviewers who can use the coding tool to extract the relevant information.

Statistical Procedures and Conventions

Calculating Effect Sizes

Where data allows, effects sizes will be computed for each study. The EPPI-Reviewer software has built-in functionality for calculating effect sizes from a range of statistics that are presented in study reports. Where necessary, other web-based resources (for example, the Campbell Collaboration's effect size calculator) and expert consultation will be used for the less common statistical representations.

For studies reporting dichotomous outcomes (for example, employment rates), both the risk ratios (RRs) and standard mean difference (SMD) effect sizes (Cohen's *d*) will be calculated. For outcomes measured on a continuous scale (for example, group differences in levels of income) we will calculate both SMDs and response ratios (RRs). By computing different effect sizes, we will be able to explore the sensitivity of the results to the selection of the effect size measure and cope with any possible loss of information arising from impossibility to compute all effect size measures from every included study. The review will correct for sample bias in the effect sizes by using the correction for sample bias procedure developed by Hedges and Olkin (1985).

Reviewers will document the computations used for the effect size estimates derived from each study. All effect sizes will be coded such that positive effect sizes represent positive outcomes (e.g., less unemployment, higher wages).

To correct for variation associated with cluster-level assignment, the unit of assignment to treatment and comparison groups will be coded for all studies, and appropriate adjustments made to effect sizes (Hedges, 2007).

We anticipate encountering the use of a continuous outcome by some authors, and a dichotomous outcome by others. For each outcome category, we will determine the number of coded effect sizes in the different metrics. Where more than one type occurs in a given outcome category, we will transform the effect size metric with the smaller proportion into the metric with the larger proportion using the Cox transform as shown by Sánchez-Meca et al. (2003). This will allow all the effect sizes for that outcome category to be analysed together.

In the event that we do not have consistency across our data (i.e., effect sizes based on either all raw data or all log-transformed data), we will consult Higgins et al. (2008) for guidance on data transformation.

Synthesis of Effect Sizes

The analysis will be conducted using the specialised built-in meta-analysis functions within EPPI-Reviewer. EPPI-Reviewer is able to support the entire meta-analysis process, from the extraction of data from primary studies to the production of forest plots and sub-group analyses.¹ It is assumed that data extracted from the studies will be pooled using meta-analysis; however, this will depend on such factors as the heterogeneity of the studies and study populations. Key features of the participants, interventions and outcomes will be described in summary tables, along with effect size estimates and methodological quality characteristics.

The review will include both randomised and quasi-experiments. We will report separate estimates of intervention effects for randomised and non-randomised studies. We will also provide separate effect sizes for each distinguishable class type of randomised and non-randomised design (even where such results are judged to be similar or are not statistically significantly different from each other). Where appropriate, we will summarise *across* designs, in addition to providing separate results *by* design. Our rationale for conducting a cross-design synthesis will be made explicit. The synthesis will separate studies with different kinds of counterfactuals (i.e., TVET versus no treatment studies will be combined separately from TVET versus some alternative intervention).

The data syntheses will be carried out using random effects statistical models. To account for differences in sample sizes for individual studies, effect sizes will (where possible) be averaged across studies by using an inverse variance weighting of the individual effect size. This weighting will result in the individual effect sizes of larger n studies being given more weight in the combined effect size. To visibly examine variability (or consistency) in the effect-size estimates, forest plots will be used to display the estimated effect sizes from each study along with their 95% confidence intervals. In the event that meta-analysis is not appropriate, forest plots will be still reported. Heterogeneity tests (Q and I²) will be used to

¹ It supports the meta-analysis of both *d* and *r* families of continuous effect size, as well as binary outcomes.

examine whether variation (or consistency) in effect-size estimates is more than would be expected (Deeks et al., 2001). The results for a random effects model and a funnel plot that plots effect size against standard error will be presented and examined for possible publication bias. Duval and Tweedie's 'trim and fill' method (2000) and/or Egger's regression test (1997) will be used to assess the impact of missing studies on the results of the meta-analysis. For the main effect, we will report a post-hoc power analysis (Borenstein et al., 2009).

To test the robustness of the results of the data synthesis, a number of sensitivity analyses will be conducted. We will examine whether the results are sensitive to:

- (i) the methodological quality of studies;
- (ii) the specific statistical procedures and methods for computing each effect size;
- (iii) our method of analysis (in particular, decisions relating to transformation between effect size metrics, the way outlier effect sizes and sample sizes were handled, and missing data imputations); and
- (iv) the way outcomes were measured and the timing at which measured.

Where appropriate, moderator analysis using meta-regression models will be used in an attempt to identify variables that are associated with larger and smaller effects for the various outcome constructs. Evidence of differential effects will be explored for trainees with different characteristics (e.g., in relation to gender, education level, work experience, length of current employment status), by programme characteristics (e.g., in relation to type of TVET, duration of training, programme location, sector/industry, and voluntariness) and by labour market conditions. We will conduct power calculations for these analyses. For each individual moderator variable, a minimum of ten studies of sufficient quality are required (Borenstein et al., 2009). If we do not have enough data points to run a meta-regression (multivariate), we will use an analogue to the ANOVA analysis (univariate) approach (described in Lipsey and Wilson, 2001).

If we have studies that are missing data that is considered essential for the review, our approach will involve: thorough attempts to contact the original investigators and funding sources; imputing the missing data with replacement values (accounting for the uncertainty by multiple imputation, adjustment to the standard error and sensitivity analysis); making explicit the methods used to impute missing data; and discussing the potential impact of missing data on the findings of the review (Higgins and Green, 2008).

Treatment of Qualitative Research

Qualitative research will not be included in this review.

REFERENCES

- ADB (2008) Indonesia: Vocational Education Strengthening Project (INVEST). Mandaluyong City, Philippines: Asian Development Bank.
- Angel-Urdinola DF, 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. Social Protection and Labor. World Bank.
- Attanasio OP, Kugler AD, 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 OP, Kugler AD, Meghir C (2009) Subsidizing vocational training for disadvantaged youth in developing countries: Evidence from a randomized trial. IZA Discussion Paper No. 4251. Bonn.
- 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.
- Balshem H, Helfand M, Schunemann HJ, Oxman AD, Kunz R, Brozek J, Vist GE, Falck-Ytter Y, Meerpohl J, Norris S, Guyatt GH (2011) GRADE guidelines: 3. Rating the quality of evidence. *Journal of Clinical Epidemiology*, 64: 401-406.
- 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. Social Protection Unit. 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: World Bank.
- Bidani B, Goh C, O'Leary C (2002) Has training helped employ Xiagang in China? A tale from two cities. Washington: World Bank.
- Bidani B, Goh C-C, Blunch N-H, 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 (2009) Evaluating job training in two Chinese cities. *Journal of Chinese Economic and Business Studies*. 7(1): 77-94.
- Borenstein M, Hedges LV, Higgins JPT, Rothstein HR (2009) Introduction to Meta-Analysis. John Wiley & Sons.

- Bouillon CP, Tejerina L (2006) Do we know what works? A systematic review of impact evaluations of social programs in Latin America and the Caribbean. Inter-American Development Bank Working Paper.
- Card D, Kluve J, Weber A (2010) Active Labour Market Policy Evaluations: A Meta-Analysis. *Economic Journal*, 120(548), F452-F477.
- Cooper H, Hedges L (eds.) (1994) *The Handbook of Research Synthesis*. New York: Russell Sage Foundation.
- Deeks JJ, Douglas AG, Bradburn MJ (2001) Statistical methods for examining heterogeneity and combining results from several studies in meta-analysis. In: Egger M, Davey Smith G; Altman DG. *Systematic Reviews in Health Care: Meta-analysis in Context*. London: BMJ Publishing Group.
- 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.
- 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.
- Fares J, Puerto OS (2009) *Towards Comprehensive Training*. Social Protection Discussion Paper. No. 0924. World Bank.
- Godfrey M (2003) *Youth Employment Policy in Developing and Transition Countries - Prevention as well as Cure*. Social Protection Discussion Paper 0320. Washington: World Bank.
- Gough D (2007) Weight of evidence: a framework for the appraisal of the quality and relevance of evidence. In J. Furlong, A. Oancea (Eds.) *Applied and Practice-based Research*. Special Edition of *Research Papers in Education*, 22, (2), 213-228.
- Greenberg D, Michalopoulos C, Robins P (2003) A Meta-Analysis of Government-Sponsored Training Programs. *Industrial and Labor Relations Review*, 57(1): 31-53.
- Hedges L (2007) Effect sizes in cluster-randomised designs. *Journal of Educational and Behavioural Statistics*, 32, 341-370.
- Hedges L, Olkin I (1985) *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.
- Hedges L, Tipton E, Johnson M (2010) Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods*, 1: 39-65.

- Hicks J, Kremer M, Mbiti I, Miguel E (2011) Vocational education voucher delivery and labor market returns: a randomized evaluation among Kenyan youth. Washington: World Bank.
- Higgins J, Thompson S, Deeks J, Altman D (2003) Measuring inconsistency in meta-analysis. *BMJ* (327):557-60.
- Higgins J, Green S (eds.) (2008) *Cochrane Handbook for Systematic Reviews of Interventions*. Chichester: Wiley-Blackwell.
- Higgins J, White I, Anzures-Cabrera J (2008) Meta-analysis of skewed data: Combining results reported on log-transformed or raw scales. *Statistics in Medicine*. 27. 6072–6092.
- Hillage J, Pollard E (1998) *Employability: developing a framework for policy analysis*. Research Report RR85. London: Department for Education and Employment.
- Ibarraran P, Rosas Shady D (2008) *Evaluating the Impact of Job Training Programs in Latin America: Evidence from IDB Funded Operations*. Inter-American Development Bank.
- ILO (2011) *Global Employment Trends for Youth: 2011 update*. Geneva: International Labour Office.
- ILO (2012) *Global Employment Trends 2012: Preventing a deeper jobs crisis*. Geneva: International Labour Office.
- IMF (2010). *Uganda: Poverty Reduction Strategy Paper*. Washington, D.C.: International Monetary Fund.
- Katz E (2008) *Programs Promoting Young Women's Employment: What Works?* Washington: World Bank.
- 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: 409-48.
- Kluve J (2006) *The Effectiveness of European Active Labor Market Policy*. IZA Discussion Paper. Bonn.
- Knowles and Behrman (2005) *The economic returns to investing in youth in developing countries: a review of the literature*. Health, Nutrition and Population (HNP) Discussion Paper. World Bank.

- Lipsey W, Wilson D (2001) *Practical Meta-analysis*. London: Sage.
- Maclean R, Wilson DN, eds. (2009) *The International Handbook of Education for the Changing World of Work: Bridging Academic and Vocational Learning*, volume 1. New York: Springer.
- Martin JP, Grubb D (2001) What works and for whom: a review of OECD countries' experiences with active labour market policies, *Swedish Economic Policy Review* 8: 9-56.
- McQuaid RW, Lindsay C (2005) The concept of employability, *Urban Studies*, 42(2): 197-219.
- Misola, NK (2010). *Improving the participation of female students in TVET programmes formerly dominated by males: the experience of selected colleges and technical schools in the Phillipines*. Bonn, Germany: UNESCO-UNEVOC.
- Palmer R et al. (2007) *Educating out of poverty? A synthesis report on Ghana, India, Kenya, Rwanda, Tanzania and South Africa*. Department for International Development. *Researching the Issues*, No 70.
- Péano S, Vergel de Dios B, Atchoaréna D, Mendosa U (2008) *Investment in Technical and Vocational Education and Training (TVET) in the Philippines*. Paris: International Institute for Educational Planning.
- Puerto OS (2007) *Interventions to Support Young Workers in Latin America and the Caribbean: Regional Report for the Youth Employment Inventory*. Washington: World Bank.
- Sánchez-Meca J, Marin-Martinez F, Chácon-Moscoso S (2003) Effect-size indices for dichotomized outcomes in meta-analysis. *Psychological Methods*, 8, 448-467.
- Spevacek A (2009) *Effectiveness of Active Labor Market Programs: A Review of Programs in Central and Eastern Europe and the Commonwealth of Independent States*
- Stavreska A (2006) *Interventions to Support Young Workers in South, East Asia and the Pacific. Regional Report for the Youth Employment Inventory*. Washington: World Bank.
- 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.
- UN (2012) *World Youth Report*. Geneva: United Nations.

- UN/DESA (2011) *World Economic Situation and Prospects 2012: Global Economic Outlook*. New York: United Nations.
- UNESCO (2010a) *Education for All Global Monitoring Report 2010. Reaching and teaching the most marginalised*. Paris: UNESCO.
- UNESCO (2010b) *Guidelines for TVET policy review*. ED/ESB/TVET/2010/02. UNESCO.
- United Nations (1992) *United Nations: Statistical charts and indicators on the situation of youth, 1970-1990*. New York.
- Van Adams A (2007) *The role of youth skills development in the transition to work: a global review*. 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.
- World Bank (2007) *Strategy to Revitalize Technical and Vocational Education and Training (TVET) in Africa*. World Bank.

APPENDIX 1: WORLD BANK LIST OF ECONOMIES (18 JULY 2011 REVISION; IN EFFECT UNTIL JULY 2012)

	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		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 Príncipe, 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

APPENDIX 2: SAMPLE SEARCH STRATEGY

The following string will be used to search ERIC and other bibliographic databases using the CSA platform (the KW function searches the descriptor field, title and abstract). It will be 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 "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

APPENDIX 3: CODING MANUAL

Section A: Administration

A.1 Name of reviewer

A.1.1 KH

A.1.2 JT

A.1.3 CB

A.1.4 MN

A.2 Linked reports

Please enter the details of other reports on this item/study (i.e., so called 'linked reports') and whether they have been used to complete this data extraction.

A.2.1 No

A.2.2 Yes (specify)

A.3 Language of main report

A.3.1 English

A.3.2 Spanish

A.3.3 Other (specify)

Section B: Study focus

B.1 What are the broad aims of the study?

B.1.1 Not stated

B.1.2 Explicitly stated (specify)

B.1.3 Implicit (specify)

B.1.4 Unclear (specify)

B.2 Do the authors report how the study was funded?

B.2.1 Not stated

B.2.2 Stated (specify)

B.2.3 Unclear (specify)

B.3 When was the study carried out?

If the authors give a year or range of years, then put that in. If not, give a 'not later than' date by looking for a date of first submission to the journal, or for clues like the publication dates of other reports from the study.

B.3.1 Not stated

B.3.2 Explicitly stated (specify)

B.3.3 Implicit (specify)

B.3.4 Unclear (specify)

B.4 Country where study conducted

B.4.1 Low-income country (specify)

B.4.2 Low-middle income country (specify)

B.4.3 Upper-middle income country (specify)

Section C: Participants

C.1 Total number of participants

If more than one group is being compared, please give numbers for each group. If numbers are given separately for male and female youth, please include both figures. NB: the sample size of the treatment and the control groups for each effect size (which may be different from the total sample size) will be collected later (when effect sizes are calculated).

C.1.1 Not stated

C.1.2 Explicitly stated (specify)

C.1.3 Implicit (specify)

C.1.4 Unclear (specify)

C.2 Age of participants

If numbers are given separately for males and females, please also report these numbers (this might require reporting numbers for different age groups for both males and females).

C.2.1 All participants within 15-24 age range

Specify average age, if reported, and range.

C.2.2 Average age within 15-24 age range

Specify average age and overall range.

C.2.3 Results reported for participants within 15-24 age range

Specify both age ranges, and any relevant average ages.

C.3 Sex of participants

C.3.1 Not stated

C.3.2 Male only

C.3.3 Female only

C.3.4 Mixed

C.3.5 Unclear (specify)

C.4 Employment status of participants at time of receipt of intervention

C.4.1 Not stated

C.4.2 In paid employment

Specify proportion in full-time work and other relevant details, if reported.

C.4.3 Mix of in paid employment and not in paid employment

Specify ns/% for the employed/unemployed, and proportion in full-time work and other relevant details, if reported.

C.4.4 Not in paid employment (all)

C.4.5 Unclear (specify)

C.5 Length of current employment/unemployment of participants

C.5.1 Not stated

C.5.2 Stated (specify)

C.5.3 Unclear (specify)

C.6 Skills/experience level of participants

C.6.1 Not stated

C.6.2 Explicitly stated (specify)

C.6.3 Implicit (specify)

C.6.4 Unclear (specify)

C.7 Is there any other useful demographic information about study participants?

Please specify any other useful demographic information about the study participants. For example, data relating to marital status, years of formal education, position in the household (whether head or not).

C.7.1 No

C.7.2 Yes (specify)

C.8 If applicable, which cohort(s) of the programme are participants drawn from?

C.8.1 Not applicable (one cohort only)

C.8.2 Not stated

C.8.3 Explicitly stated (specify)

C.8.4 Implicit (specify)

C.8.5 Unclear (specify)

Section D: Intervention

D.1 Formal name

D.1.1 Not stated

D.1.2 Stated (specify)

D.1.3 Unclear (specify)

D.2 Aim(s)/rationale of intervention

D.2.1 Not stated

D.2.2 Explicitly stated (specify)

D.2.3 Implicit (specify)

D.2.4 Unclear (specify)

D.3 Theory/hypothesis for how intervention should work

D.3.1 Not stated

D.3.2 Stated (specify)

D.3.3 Unclear (specify)

D.4 Year intervention started (and ended, if relevant)

Date of initiation. Please indicate if still ongoing.

D.4.1 Not stated

D.4.2 Stated (specify)

D.4.3 Unclear (specify)

D.5 Single- vs. multi-component

Is the 'treatment' condition a single activity or a combination of activities? For example, does the intervention combine training with wage subsidies?

D.5.1 Not stated

D.5.2 Single-component/service

D.5.3 Multi-component/service

D.5.4 Unclear (specify)

D.6 Mode(s) of TVET

Select all that apply. Refer to definitions in protocol. If necessary, provide brief details to justify your choice.

- D.6.1 Not stated
- D.6.2 Technical education
- D.6.3 Vocational education
- D.6.4 Vocational training
- D.6.5 Apprenticeship
- D.6.6 On-the-job training
- D.6.7 Other (specify)
- D.6.8 Unclear (specify)

D.7a Programme coverage

- D.7a.1 Not stated
- D.7a.2 National
- D.7a.3 Regional (specify)
- D.7a.4 Local (specify)
- D.7a.5 Unclear (specify)

D.7b Programme location

Select all that apply.

- D.7b.1 Not stated
- D.7b.2 Urban (specify)
- D.7b.3 Rural (specify)
- D.7b.4 Unclear (specify)

D.8 Political/economic/social/cultural context (at time of intervention)

Note any relevant contextual factors: i.e., the economic/macro-economic/political/social/cultural environment of the country at time of the intervention. Consider the following questions. Were there any political barriers to implementing this intervention? Were there any social/cultural norms which prevented acceptance of the intervention? What is the prevalence of the problem of interest (e.g., what is the level of youth unemployment at the time). Describe all relevant labour market conditions reported in the paper. Detail any relevant information about level of labour market regulation (in terms of employment protection and minimum wages) if provided.

- D.8.1 Not stated
- D.8.2 Stated (specify)
- D.8.3 Unclear (specify)

D.9 Target group

Select all that apply.

- D.9.1 Not stated
- D.9.2 Young people
- D.9.3 Un/underemployed
- D.9.4 Low educational attainment
- D.9.5 Low income/'poor'

- D.9.6 No/little work experience
- D.9.7 Low/unskilled
- D.9.8 Low SES/'disadvantaged' (i.e., broader than low income)
- D.9.9 Other (specify)
- D.9.10 Unclear (specify)

D.10 Voluntary vs. mandatory

- D.10.1 Not stated
- D.10.2 Voluntary
- D.10.3 Mandatory
- D.10.4 Unclear (specify)

D.11 Is the intervention focused on specific sectors/industries?

- D.11.1 Not stated
- D.11.2 Yes (specify)
- D.11.3 No
- D.11.4 Unclear (specify)

D.12 Institutional setting

- D.12.1 Not stated
- D.12.2 Schools
- D.12.3 Colleges
- D.12.4 Training centres
- D.12.5 Worksites
- D.12.6 Other (specify)
- D.12.7 Unclear (specify)

D.13 Type of learning arrangement

Formal learning consists of learning that occurs within an organised and structured context (formal education, in-company training), and that is designed as learning. It may lead to a formal recognition (diploma, certificate). Formal learning is intentional from the learner's perspective. Non-formal learning consists of learning embedded in planned activities that are not explicitly designated as learning, but which contain an important learning element. Non-formal learning is intentional from the learner's point of view.

- D.13.1 Not stated
- D.13.2 Formal
- D.13.3 Non-formal
- D.13.4 Unclear (specify)

D.14 Mode of delivery

- D.14.1 Not stated
- D.14.2 Face-to-face
- D.14.3 Online
- D.14.4 Distance learning
- D.14.5 Unclear (specify)

D.15 Content/substance of the intervention

Provide details about the specific services that were provide as part of the intervention. Whenever possible copy the authors' description from the report word for word.

D.15.1 Not stated

D.15.2 Stated (specify)

D.15.3 Unclear (specify)

D.16 Employer involvement

For example: involvement in provision of training as part of TVET programme; deciding on the context of the TVET programme; deciding on the mix of provision between different programmes; availability of data on labour market outcomes for prospective students.

D.16.1 Not stated

D.16.2 Stated (specify)

D.16.3 Unclear (specify)

D.17 Who delivered/provided the intervention?

This question refers to: (a) the organisation who is responsible for providing the intervention. If any relevant information (e.g., about the structure of the organisation) is provided in the study report, state it here. (b) the actual personnel who taught the participants (i.e., were in personal contact with them). Provide as much information as possible. Important to note whether the providers have the necessary skills to deliver this intervention. Have they been specifically trained? Note any relevant qualifications, etc.

D.17.1 Not stated

D.17.2 Stated (specify)

D.17.3 Unclear (specify)

D.18 Duration of the intervention (for participants in the study)

For how long did they receive 'treatment'? Please provide start and end dates, if information is provided in paper.

D.18.1 Not stated

D.18.2 Stated (specify)

D.18.3 Unclear (specify)

D.19 Frequency/intensity of the intervention

D.19.1 Not stated

D.19.2 Stated (specify)

D.19.3 Unclear (specify)

D.20 Is the course/programme situated within a quality assurance framework?

Is this training part of a system that has a standards framework and quality assurance processes? If applicable, state whether international or regional standards.

D.20.1 Yes (specify)

D.20.2. No/not stated (specify)

D.21 Who designed/developed the intervention?

This may be the same organisation (e.g., an NGO) as the funder/sponsor of the intervention. You may be unable to distinguish between design and develop.

D.21.1 Not stated

D.21.2 Stated (specify)

D.21.3 Unclear (specify)

D.22 Funding source for intervention

D.22.1 Not stated

D.22.2 Stated (specify)

D.22.3 Unclear (specify)

D.23 Relationship of study funder/sponsor to intervention

D.23.1 Not stated

D.23.2 Stated (specify)

D.23.3 Unclear (specify)

D.24 Relationship between evaluator and intervention

D.24.1 Not stated

D.24.2 Stated (specify)

D.24.3 Unclear (specify)

D.25 Is there any other relevant information about the intervention?

For example (a) were there any additional/more specific eligibility criteria (i.e., in addition to the target group responses listed in question D.9); (b) was the training based on the labour demand of specific companies (i.e., were young people being trained in areas in which there was already a market demand)? [NB: this may have already been discussed as part of question D.11]

D.25.1 Details

D.26 What treatment/intervention did the control/comparison group receive?

If specified in the report, describe in detail what the control/comparison group(s) were exposed to.

D.26.1 Not stated

D.26.2 No treatment

D.26.3 Treatment as usual (specify)

D.26.4 Alternative intervention (specify)

D.26.5 Other (specify)

D.26.6 Unclear (specify)

Section E: Methods

E.1a Study timing

There are a range of accepted approaches to determining an appropriate comparison group for counterfactual analysis, using either prospective (ex ante) or retrospective (ex post) evaluation design. Prospective evaluations begin during the design phase of the intervention, involving collection of baseline and end-line data from intervention beneficiaries (the 'treatment group') and non-beneficiaries (the 'comparison group'), and may also involve selection of individuals or communities into treatment and comparison groups. Retrospective evaluations are usually conducted after the implementation phase, and may exploit existing survey data, although the best evaluations will collect data as

close to baseline as possible, to ensure comparability of intervention and comparison groups.

- E.1a.1 Prospective (ex-ante) evaluation design
- E.1a.2 Retrospective (ex-post) evaluation design
- E.1a.3 Unclear (specify)

E.1b Design

- E.1b.1 Randomised experiment
Include group randomised trials here if number of aggregates is adequate and if properly analysed.
- E.1b.2 Randomised experiment with units of analysis discrepancy or very small number of aggregate units
For example, classrooms randomly assigned to conditions, but individuals treated as unit of analysis; or one school per condition.
- E.1b.3 Quasi-experiment: interrupted time series
- E.1b.4 Quasi-experiment: regression discontinuity
- E.1b.5 Quasi-experiment: non-equivalent comparison group
Methods of analysis may have been used to control for differences between groups.
- E.1b.6 Case-control
- E.1b.7 Within-group comparison (i.e., pre-test-post-test)
- E.1b.8 Other
For example, design that has both random assignment and self-selection into several groups.

E.2a Method of group assignment

This question is about assignment of subjects to the intervention and control conditions. All studies being coded have included a counterfactual population. How is the counterfactual defined? Errors in this process of allocating individuals to groups cause selection bias (i.e., unequal distribution of potential confounding factors). This question focuses on the initial method of assignment to groups, regardless of subsequent degradations due to attrition, refusal, etc, prior to treatment onset. These latter problems are coded elsewhere.

- E.2a.1 Prospective allocation into more than one group
- E.2a.2 Use of pre-existing differences to create comparison groups
- E.2a.3 Other (specify)
- E.2a.4 Unclear (specify)

E.2b Method of group assignment

This question is about assignment of subjects to the intervention and control conditions. This question focuses on the initial method of assignment to groups, regardless of subsequent degradations due to attrition, refusal, etc, prior to treatment onset. These latter problems are coded elsewhere.

- E.2b.1 Randomly (after matching, stratification, blocking)
The entire sample is matched/stratified/blocked first, then assigned to treatment and comparison groups within pairs or blocks. This

does not refer to blocking after treatment for the data analysis. In any randomised trial it is desirable that the comparison groups should be as similar as possible as regards those characteristics that might influence the response to the intervention. Stratified randomisation is used to ensure that equal numbers of participants with a characteristic thought to affect prognosis or response to the intervention will be allocated to each comparison group. Stratified randomisation is performed either by performing separate randomisation (often using random permuted blocks) for each strata, or by using minimisation.

E.2b.2 Randomly (without matching, stratification, blocking)

This also includes cases when every other person goes to the control group.

E.2b.3 Regression discontinuity

Quantitative cutting point defines groups on some continuum.

E.2b.4 Quasi-random procedure (after matching etc)

Presumed to produce comparable groups (no obvious differences).

This applies to groups which have individuals apparently randomly assigned by some naturally occurring process, e.g. next person to walk in the door. The key here is that the procedure used to select groups doesn't involve individual characteristics of persons so that the groups generated should be essentially equivalent.

E.2b.5 Quasi-random procedure (with matching etc)

E.2b.6 Non-random, but matched or statistically controlled (ONLY on pre-test measures of some or all variables used later as outcome measures)

Matching refers to the process by which comparison groups are generated by identifying individuals or groups that are comparable to the treatment group using various characteristics of the treatment group (matching on observables, or on 'propensity scores').

Statistical control refers to inclusion of the matching variable as a covariate in an ANCOVA or multiple regression analysis. Matching can be done individually, e.g., by selecting a control subject for each intervention subject who is the same age, gender, and so forth, or on a group basis, e.g., by selecting comparison schools that have the same demographic makeup and academic profile of treatment schools. Similarly, statistical control variables can be used at either the individual or school level. Propensity score matching (PSM) identifies a group of individuals, households or firms with the same observable characteristics as those participating in the project. It does this by estimating a statistical model of the probability of participating (propensity to participate) using a regression model

with participation as the zero-one dependent variable, and a set of observable characteristics, which must be unaffected by the intervention, as the explanatory variables. The coefficients are used to calculate a propensity score, and participants matched with non-participants based on having similar propensity scores. In practice there are a range of ways of performing this matching, with the most common being to match each participant with their five 'nearest neighbour' non-participants (i.e. the five non-participants with the closest propensity score). Propensity score matching can be attractive for two reasons. First, comparison group data may have been collected but are thought not to be representative because of selection bias. Second, there may be data only on the treatment group but not the control. A different, possibly nationwide, data set can then be used to construct a comparison group using PSM.

E.2b.7 Non-random, but matched or statistically controlled (on pre-test measures AND other personal characteristics, such as demographics)

See E.2b.6

E.2b.8 Non-random, but matched or statistically controlled (ONLY on demographic variables)

See E.2b.6

E.2b.9 Non-random, not matched, but pre-treatment equivalence information is available regarding the nature of the group differences

E.2b.10 Unclear

E.3 Which method was used to generate the allocation sequence?

Provide details of the method used to generate the allocation sequence, including details of any restrictions (e.g., blocking, stratification).

E.3.1 Not applicable (no prospective allocation)

E.3.2 Not stated

E.3.3 Random (specify method)

A method that uses the play of chance to assign participants to comparison groups in a trial (e.g. by using a random numbers table; computer-generated random sequence; shuffling cards or envelopes; throwing dice; drawing of lots; computer random number generator; coin tossing; minimization. Random allocation implies that each individual or unit being entered into a trial has the same chance of receiving each of the possible interventions.

E.3.4 Quasi-random (specify method)

Presumed to produce comparable groups (no obvious differences). This applies to groups which have individuals apparently randomly assigned by some naturally occurring process, e.g. next person to walk in the door. The key here is that the procedure used to select

groups doesn't involve individual characteristics of persons so that the groups generated should be essentially equivalent.

E.3.5 Non-random (specify method)

For example, assignment based on date of birth, case record number, date of presentation, allocation by availability of the intervention, according to the preference of the study subject, training provider or investigator, or based on the result of a test. (NB: some researchers will report methods such as use of date of birth as quasi-random or even random).

E.3.6 Unclear (specify)

E.4 Was the allocation sequence concealed?

Does the paper refer to any information about whether or not the allocation was concealed (from participants and/or researchers/others creating the groups)? For example, is there reference to sequentially numbered sealed envelopes or 'centralised' allocation (eg web-based randomisation)?

E.4.1 Not applicable (no prospective allocation)

E.4.2 Not stated

E.4.3 Yes (specify)

E.4.4 No (specify)

E.4.5 Unclear (specify)

Select if method of concealment is not described in sufficient details to allow a definite judgment: for example, it is reported that envelopes were used, but it is not stated if they were sequentially numbered, opaque and sealed.

E.5 What was the unit of allocation/assignment?

For interventions that naturally occur in groups of individuals, random allocation of participants may be inappropriate. In these cases, the unit of random allocation may be the group or cluster, rather than the individual. The paper will therefore refer to both the intervention and control groups being comprised of a number of clusters rather than a number of individuals. In making your selection, indicate all that apply and give further details where possible.

E.5.1 Not applicable (no prospective allocation)

E.5.2 Not stated

E.5.3 Individuals

E.5.4 Groupings (clusters) of individuals

E.5.5 Unclear (specify)

E.6 Number of groups

This is the total number of groups allocated to the intervention and control conditions.

E.6.1 Two (intervention and control)

E.6.2 Three (specify)

E.6.3 Four or more (specify)

E.6.4 Unclear (specify)

E.7 When were the measurements of the outcome variables made, in relation to the intervention?

If at least one of the outcome variables is measured before and after the intervention, use the 'before and after' category.

E.7.1 Before and after

Select if pre-intervention measures of outcomes and other important variables collected at baseline and incorporated into the analysis.

E.7.2 Only after

E.7.3 Unclear (specify)

E.8 Format of impact measurements

Are the impact variables in cross-sectional or longitudinal format? What we are looking for here is the use of time-series (broadly-defined) measurements taken after (and possibly before) the intervention. NB: 'before and after' studies are not necessarily longitudinal.

E.8.1 Cross-sectional

E.8.2 Longitudinal/ time-series measurements

E.8.3 Unclear (specify)

E.9 Number and timing of impact/outcome measurements

Indicate exactly when the post-treatment measurements were taken (in relation to the end of the training period)

E.9.1 One (specify)

One 'post-test measurement' occurring at or after programme completion.

E.9.2 Two (specify)

One 'post-test measurement' occurring at or after programme completion, followed by one 'follow-up'.

E.9.3 Three or more (specify)

One 'post-test measurement' occurring at or after programme completion, followed by two 'follow-ups'.

E.9.4 Unclear (specify)

E.10 Study design summary

Describe the study design in your own words, drawing upon and elaborating the answers you have already given (questions E1-E9). The emphasis is on making clear the basic strategies that the authors have used to improve upon the single group before and after design (for example, use of control group, taking more measurements before and after the intervention implementation). Most of these strategies are dealt with in the previous questions. Other design elements to look for are: staggering the introduction of the intervention among groups; adding a reversal to the intervention; and using additional outcome measures. The focus here is on the actual design, rather than how well it was executed. Attention should be paid to the design 'features' of the primary studies (such as how participants were allocated to groups, which parts of the study were prospective, etc.) rather than design 'labels' (such as 'cohort' or 'cross-sectional') as the risk of bias is

influenced by the specific features of a study rather than a broad categorisation of the approach taken.

E.10.1 Experiment (specify)

E.10.2 Quasi-experiment (specify)

E.11 Details of sampling procedure

Describe the process/procedure undertaken to create the study sample, including details about the sampling frame. Here the focus is on whether subjects were chosen randomly from (i.e., were representative of) the target population (in addition to random assignment to groups, if relevant). It may be the case that the sample has been drawn randomly, but not from the target population (e.g., drawn randomly from a list of (highly motivated) people who applied to receive the training). Please note if the authors refer to self-selection, pre-screening of participants, discounting participants that did not run to completion and excluding subjects who have recently moved into or out of the study area. Also consider whether it is likely that participants (from either group) received an unintended intervention (contamination or co-intervention) that may influence the results. Was the control group selected in such a way as to minimise the chances of spillover and crossover effects?

E.11.1 Details

E.12 Planned sample size

Give details for each group separately

E.12.1 Not applicable (specify)

E.12.2 Not stated

E.12.3 Stated (specify)

E.12.4 Unclear (specify)

E.13 Details of methods used to recruit people into the study

For example, letters of invitation, face-to-face contact. Include information about whether or not the consent of participants (or other consent) was sought. Consider whether methods of recruitment were likely to introduce bias in the selection of the sample.

E.13.1 Not applicable (specify)

E.13.2 Not stated

E.13.3 Explicitly stated (specify)

E.13.4 Implicit (specify)

E.13.5 Unclear (specify)

E.14 What percentage of selected individuals agreed to participate?

E.14.1 Not applicable (specify)

E.14.2 Not stated

E.14.3 80-100% (specify)

E.14.4 60-79% (specify)

E.14.5 Less than 60% (specify)

E.14.6 Unclear (specify)

E.15 Comparability of groups (at baseline)

Provide details about the comparability of groups at baseline. The next question will

capture information about any differences that arise from individuals dropping out of the study. Please make a note of any descriptive and/or statistical comparisons. The following are examples of confounders (i.e., variables on which groups may have been compared): race, age, sex, marital status, SES, education, pre-intervention score on outcome measure. Please indicate (i) the number of variables on which treatment and comparison group differences were statistically compared, (ii) where any differences are reported as being statistically significant, and (iii) whether you judge differences to be important or unimportant. NB: an 'important' difference means a difference on several variables relevant to the outcome variables, or on a major variable (e.g., SES), or large differences, prior to the intervention. A statistical comparison is one in which a statistical test was performed by the authors, whether they provide data or not (e.g., 'no statistically significant differences were found'). For ex-post studies where statistical techniques have been used to create equivalent groups, this may have been done post test (i.e., no pre-test comparisons) - if so, make a note of this.

E.15.1 Not stated

E.15.2 Explicitly stated (specify)

E.15.3 Implicit (specify)

E.15.4 Unclear (specify)

E.16 If the study involves studying samples prospectively over time, what proportion of the sample dropped out over the course of the study?

If the study involves more than one group, give drop-out rates for each group separately.

E.16.1 Not applicable (not following samples prospectively over time)

E.16.2 Not stated

E.16.3 Explicitly stated (specify)

E.16.4 Implicit (specify)

E.16.5 Unclear (specify)

E.17 Do the authors provide any information on whether, and/or how, those who dropped out of the study differ from those who remained in the study?

E.17.1 Not applicable (not following samples prospectively over time)

E.17.2 Not applicable (no drop outs)

E.17.3 Yes (specify)

E.17.4 No

E.18 Details of data collection (methods, tools and personnel)

Provide details, including names (if relevant) for tools used to collect data. If relevant, state whether source of any tool is cited in the paper (or there is reference to another paper where the instrument is detailed in full). Provide details about those who collected the data (e.g., whether it was the authors themselves or others who collected data). The nature of the actual data that were collected should be clear from your answer to this question. If different tools/methods used for baseline and outcome measurement, ensure your answers specifies this clearly. If more than one tool/method is used in measurements of outcomes, include all relevant details. It is important to capture here whether the study uses different methods to measure the same variable (i.e., to avoid mono-method bias).

- E.18.1 Not stated
- E.18.2 Explicitly stated (specify)
- E.18.3 Implicit (specify)
- E.18.4 Unclear (specify)

E.19 Was there concealment of which groups individuals were assigned to, and/or concealment of other key factors, from those collecting outcome data?

Is there any information relating to whether the intended blinding was effective?

- E.19.1 Not applicable (analysis of existing data)
- E.19.2 Not stated
- E.19.3 Yes (specify)
- E.19.4 No
- E.19.5 Unclear (specify)

E.20 Were the groups treated equally in other respects?

Consider each of the following three issues: (a) Were the data collection methods/measures for the intervention and control groups the same (i.e., was the same survey used)? (b) Were the data collection settings the same for both the intervention and control groups? (If there was more than one treatment group, were the settings the same for different groups of subjects?)

- E.20.1 Not stated
- E.20.2 Yes (specify)
- E.20.3 No (specify)
- E.20.4 Unclear (specify)

E.21 Do the authors' describe any ways they addressed the repeatability/reliability of their data collection tools/methods?

Are any test-retest methods described? Did the authors look at inter-rater reliability, or re-test a sample of results to see if they got the same answer? Where more than one tool was employed, provide details for each.

- E.21.1 Yes (specify)
- E.21.2 No

E.22 Do the authors describe any ways they have addressed the validity/trustworthiness of their data collection tools/methods?

Do the authors report any previous validation of the tools, published versions of the tools, involvement of target population in the development of the tools? Where more than one tool was employed, please provide details for each. Were different sources of data used? Was any triangulation of data carried out: for example self-reported attainment matched to official records?

- E.22.1 Yes (specify)
- E.22.2 No

E.23 Details of data analysis (methods, tools and personnel)

Provide full details of the estimation process/model specification/statistical analyses. If more than one method of estimation is used, make a note of all. Where appropriate, refer to your answers to other questions.

- E.23.1 Not stated
- E.23.2 Explicitly stated (specify)
- E.23.3 Implicit (specify)
- E.23.4 Unclear (specify)

E.24 Do the authors describe strategies to minimise bias from confounding variables?

Often variables other than the intervention may account for the reported outcomes. The degree to which confounds are accounted for affects the strength of causal inference. Please indicate the method used to control for confounders at the analysis stage (e.g., regression modelling with propensity scores or covariates) and the percentage of relevant confounders that were controlled (noting relevant ones not controlled). Where appropriate, refer to your answers to other questions.

- E.24.1 Not applicable (robust randomised experiment)
Select this answer only if E.3.3 and E.4.3 have been selected.
- E.24.2 Yes (specify)
- E.24.3 No
- E.24.4 Unclear (specify)

E.25 Do authors describe any strategies used to address attrition or other forms of missing data?

Study results can be biased by participant attrition and other forms of missing data. Statistical methods as supported by theory and research can be employed to control for missing data and attrition that would bias results, but studies with no attrition or missing data needing adjustment provide the strongest evidence that results are not biased. Where appropriate, refer to your answers to other questions.

- E.25.1 No attrition or missing data needing adjustment
- E.25.2 Missing data and attrition were taken into account inadequately, or there was too much to control for bias
Taking an average of the whole group and using this as the basis of the imputed data would be judged inadequate.
- E.25.3 Taken into account by simple estimates of data and observations, or by demonstrations of similarity between remaining participants and those lost to attrition
- E.25.4 Taken into account by more sophisticated methods that model missing data, observations, or participants
- E.25.5 Unclear (specify)

E.26a: Model assumptions: do authors assume/report that there is normal distribution/outliers (or correct for this)?

- E.26a.1 Not applicable (no statistical modelling)
- E.26a.2 Yes (specify)
- E.26a.3 No/not stated (specify)

E.26b: Model assumptions (independence of predictor variables): do the authors test for multicollinearity (and, if relevant, correct for it)

Multicollinearity is a statistical phenomenon in which two or more predictor variables in a multiple regression model are highly correlated. In this situation the coefficient estimates may change erratically in response to small changes in the model or the data)

E.26b.1 Not applicable (no statistical modelling)

E.26b.2 Yes (specify)

E.26b.3 No/not stated (specify)

E.26b.4 Unclear (specify)

Score 'Unclear' if the results of any test are not reported.

E.26c: Model assumptions: is a random sample used?

Answer 'yes' if reported that random and there is <10% attrition or if response rate >80%.

E.26c.1 Not applicable (no statistical modelling)

E.26c.2 Yes specify)

E.26c.3 No/not stated (specify)

E.26c.4 Unclear (specify)

Select 'unclear' if reported that random but no further details are provided.

E.26d: Model assumptions (linearity): do the authors test for linearity (and, if relevant, correct for this)?

If you fit a linear model to data which are nonlinearly related, your predictions are likely to be seriously in error, especially when you extrapolate beyond the range of the sample data. How to detect: nonlinearity is usually most evident in a plot of the observed versus predicted values or a plot of residuals versus predicted values, which are a part of standard regression output.

E.26d.1 Not applicable (no statistical modelling)

E.26d.2 Yes (specify)

E.26d.3 No/not stated (specify)

E.26e Model assumptions (homoscedasticity): do the authors test for homoscedasticity (and, if relevant, correct for this)

Homoscedasticity is where the variance of the error is constant across observations. For studies using OLS regression methods, report whether the authors test for homoscedasticity and non-normality in the error distribution (e.g. through a Breusch-Pagan test for heteroscedasticity or through Kolmogorov-Smirnov test for non-normality) and if relevant correct for them (e.g. use of log transformation in the dependent variable).

E.26e.1 Not applicable (no statistical modelling)

E.26e.2 Yes (specify)

E.26e.3 No/not stated (specify)

E.27 Analysis of model goodness of fit and statistical analysis of the estimated parameters

Once a regression model has been constructed, it may be important to confirm the goodness of fit of the model and the statistical significance of the estimated parameters. Commonly used checks of goodness of fit include the R-squared, graphical analyses of the pattern of residuals (e.g., scatter plots; normal probability plots; run charts) and hypothesis testing. Statistical significance can be checked by an F-test of the overall fit,

followed by t-tests of individual parameters. Interpretations of these diagnostic tests rest heavily on the model assumptions. Although examination of the residuals can be used to invalidate a model, the results of a t-test or F-test are sometimes more difficult to interpret if the model's assumptions are violated.

E.27.1 Adjusted R-squared is reported

E.27.2 Graphical analysis of residuals

E.27.3 F test for overall model fit

E.27.4 T tests of individual parameters

E.27.5 Other (specify)

E.28 Do the authors describe any further specification checks and/or sensitivity/robustness analyses?

Model validation is an important step in the model building sequence. A sensitivity analysis might involve the researchers looking at the effect of varying the inputs of a model on the output (i.e., changing one factor at a time, to see what effect this produces on the output). Robustness may have been tested by assessing the model against a set of data that was not used to create the model. It may have been reported in the paper that they executed the model repeatedly. Monte Carlo filtering may have been used if the researchers were interested in which factors are most responsible for generating high/low values of the output. Generalized score and Wald tests test certain aspects of the specification of panel probit models. Rosenbaum has developed a method of sensitivity analysis (for matched data) to assess if one's estimated based on matching is robust to the possible presence of an unobserved confounder, the key assumption for matching based analyses. His sensitivity analysis for matched data provides a specific statement about the magnitude of hidden bias that would need to be present to explain the associations actually observed. The Hausman test is a statistical hypothesis test that evaluates if a statistical model corresponds to the data.

E.29.1 Not applicable (specify)

E.29.2 Yes (specify)

E.29.3 No

E.29.4 Unclear (specify)

E.29 On what basis was data analysis carried out?

Impact evaluations which compare average outcomes in the treatment group, irrespective of beneficiary participation (also referred to as 'compliance' or 'adherence'), to outcomes in the comparison group are referred to as 'intention to treat' (or intervene) (ITT) analyses. Impact evaluations which compare outcomes among beneficiaries who comply or adhere to the intervention in the treatment group to outcomes in the control group are referred to as 'intervention received' or treatment-on-the-treated (TOT) analyses. ITT therefore provides a lower-bound estimate of impact, but is arguably of greater policy relevance than TOT in the analysis of voluntary programmes.

E.29.1 Not applicable (no prospective allocation)

E.29.2 Not stated

E.29.3 Intention to treat (or intervene)

E.29.4 Intervention received

E.29.5 Unclear (specify)

E.30 Any other issues relating to the study design or methods

E.30.1 Details

SOURCES OF SUPPORT

Financial support for this review is provided by the Australian Agency for International Development (AusAID).

DECLARATIONS OF INTEREST

The review authors have no conflicts of interest to report.

REVIEW TEAM

Lead reviewer:

Name: Janice Tripney

Title: Research Officer

Affiliation: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London

Address: 18 Woburn Square

City, State, Province or County: London

Postal Code: WC1H 0NR

Country: UK

Phone: 020 7911 5334

Email: j.tripney@ioe.ac.uk

Co-author:

Name: Dr Mark Newman

Title: Reader in Evidence-informed Policy and Practice in Education and Social Policy

Affiliation: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London

Address: 18 Woburn Square

City, State, Province or County: London

Postal Code: WC1H 0NR

Country: UK

Phone: 020 7612 6575

Email: m.newman@ioe.ac.uk

Co-author:

Name: Kimberly Hovish

Title: Research Officer

Affiliation: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London

Address: 18 Woburn Square

City, State, Province or County: London

Postal Code: WC1H 0NR

Country: UK

Email: k.hovish@ioe.ac.uk

Co-author:

Name: Dr. Chris Brown

Title: Research Officer

Affiliation: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London

Address: 18 Woburn Square

City, State, Province or County: London

Postal Code: WC1H 0NR

Country: UK

Email: christopher.brown@mac.com

ROLES AND RESPONSIBILITIES

Please give brief description of content and methodological expertise within the review team. The recommended optimal review team composition includes at least one person on the review team who has content expertise, at least one person who has methodological expertise and at least one person who has statistical expertise. It is also recommended to have one person with information retrieval expertise.

Who is responsible for the below areas? Please list their names:

- Content: Janice Tripney, Mark Newman
- Systematic review methods: Janice Tripney, Mark Newman
- Statistical analysis: Mark Newman, Janice Tripney
- Information retrieval: Janice Tripney

PRELIMINARY TIMEFRAME

The team will commence working on the systematic review upon approval by the Campbell Collaboration editorial staff. We plan to complete the review by June 2012.

PLANS FOR UPDATING THE REVIEW

Janice Tripney is responsible for maintaining the review in light of new evidence, comments and criticisms, and ensuring that the review is updated (at three-yearly intervals).

AUTHORS' RESPONSIBILITIES

By completing this form, you accept responsibility for preparing, maintaining and updating the review in accordance with Campbell Collaboration policy. The Campbell Collaboration will provide as much support as possible to assist with the preparation of the review.

A draft review must be submitted to the relevant Coordinating Group within two years of protocol publication. If drafts are not submitted before the agreed deadlines, or if we are unable to contact you for an extended period, the relevant Coordinating Group has the right to de-register the title or transfer the title to alternative authors. The Coordinating Group also has the right to de-register or transfer the title if it does not meet the standards of the Coordinating Group and/or the Campbell Collaboration.

PUBLICATION IN THE CAMPBELL LIBRARY

The support of the Campbell Collaboration and the relevant Coordinating Group in preparing your review is conditional upon your agreement to publish the protocol, finished review and subsequent updates in the Campbell Library. Concurrent publication in other journals is encouraged. However, a Campbell systematic review should be published either before, or at the same time as, its publication in other journals. Authors should not publish Campbell reviews in journals before they are ready for publication in the Campbell Library. Authors should remember to include the statement: "This is a version of a Campbell review, which is available in The Campbell Library" when publishing in journals or other venues.

I understand the commitment required to undertake a Campbell review, and agree to publish in the Campbell Library. Signed on behalf of the authors:

Form completed by: Janice Tripney

Date: 13 April 2012