

EN

The value of learning: evaluation and impact of education and training

Third report on vocational
training research in Europe
Synthesis report



The value of learning

Evaluation and impact of education and training

Third report on vocational
training research in Europe:
synthesis report

Pascaline Descy
Manfred Tessaring

A great deal of additional information on the European Union is available on the Internet.
It can be accessed through the Europa server (<http://europa.eu.int>).

Cataloguing data can be found at the end of this publication.

Luxembourg: Office for Official Publications of the European Communities, 2005

ISBN 92-896-0336-4

© European Centre for the Development of Vocational Training, 2005
All rights reserved.

*Designed by Colibri Ltd. – Greece
Printed in Belgium*

The **European Centre for the Development of Vocational Training** (Cedefop) is the European Union's reference centre for vocational education and training.

We provide information on and analyses of vocational education and training systems, policies, research and practice.

Cedefop was established in 1975 by Council Regulation (EEC) No 337/75.

Europe 123
GR-570 01 Thessaloniki (Pylea)

Postal address:
PO Box 22427
GR-551 02 Thessaloniki

Tel. (30) 23 10 49 01 11

Fax (30) 23 10 49 00 20

E-mail: info@cedefop.eu.int

Homepage: www.cedefop.eu.int

Interactive website: www.trainingvillage.gr

Acknowledgements:

The authors and Cedefop should like to thank all those who have contributed to this report:

- the authors of the background reports for their work and professionalism;
- Søren Kristensen for his input and active contribution throughout the preparation of this report;
- Sarah Elson Rogers, Éric Fries Guggenheim, Barry Nyhan and Dóra Stefánsdóttir for their active contributions, reflections and peer-reviewing of the various parts;
- Phillipe Tissot and Alena Zukersteinová for preparing the glossary of key terms related to evaluation and impact of education and training;
- Béatrice Herpin, Marena Zoppi and Silvia Del Panta for their hard work and dedication ensuring the secretarial support of the project;
- Cedefop editing, translation and publication services.

Edited by: Cedefop

Pascaline Descy, Manfred Tessaring. *Project managers*

Published under the responsibility of:

Johan van Rens, *Director*

Stavros Stavrou, *Deputy Director*

Table of contents

Abbreviations	2
Contributions to the background report of the third research report	5
Impact of education and training	6
The foundations of evaluation and impact research	6
Evaluation of systems and programmes	6
Preface by Johan van Rens	7
Introduction	8
Part 1 Understanding evaluation	19
1. Introduction to evaluation approaches and functions	21
2. Philosophies and theories of evaluation	28
3. Evaluation traditions and cultures	36
4. Standards for evaluation	42
Part 2 Approaches and methods of evaluation and impact research	47
1. Programme evaluation approaches and methods	51
2. Impact research: approaches and methods	74
Part 3 Evaluation of education and training in a changing European context	103
1. Changing contexts and drivers for change in education and training	108
2. Principles, models and practices for evaluating education and training	122
3. Policy evaluation and policy learning in action	142
4. Combating labour market exclusion: does training work?	168
Part 4 Impact and benefits of education and training	191
1. The determinants of economic growth at macro and regional level	196
2. Non-material benefits and externalities of education and training	214
3. The significance of education and training for company performance	227
4. Individual benefits of education and training	247
Annexes	
1. Key terms related to evaluation and impact of education and training	271
2. Comparing microeconomic methods of evaluation: a numeric example	274
3. Changing contexts and drivers for change for education and training: detailed tables and figures	277
4. Selected results and recommendations from the OECD thematic review on adult learning	286
5. Evaluation associations	289
6. Journals on evaluation in Europe and beyond: alphabetical list	291
References	295

Abbreviations

Country abbreviations

A	Austria
AU	Australia
B	Belgium
CA	Canada
CH	Switzerland
CY	Cyprus
CZ	Czech Republic
D	Germany
DK	Denmark
E	Spain
EE	Estonia
EL	Greece
F	France
FIN	Finland
HU	Hungary
I	Italy
IRL	Ireland
IS	Iceland
L	Luxembourg
LT	Lithuania
LV	Latvia
MT	Malta
NL	Netherlands
NO	Norway
P	Portugal
PL	Poland
S	Sweden
SI	Slovenia
SK	Slovakia
UK	United Kingdom
US	United States of America

Frequently used abbreviations

ACC	Accession countries ⁽¹⁾
ALM	Active labour market
ALMP	Active labour-market policy
ASTD	American Society of Training and Development
BCS70	1970 British Cohort Study
BHPS	British household panel survey
CBA	Cost-benefit analysis (method)
CEEC	Central and eastern European countries
CVT	Continuing vocational training
CVTS	Continuing vocational training survey
DCF	Discounted cash flow
Duoqual	Dual qualifications and vocational mobility (project)
ECHP	European Community household panel
EEPs	Employment enhancement programmes
EES	European employment strategy
ESF	European Social Fund
ETF	European Training Foundation
EVS	European value survey
FAME	Vocational identity, flexibility and mobility in the European labour market (project)
GCSE	General compulsory secondary education
GDP	Gross domestic product
GLHS	German life history study
Globalife	Life courses in the globalisation process (project)
GNP	Gross national product

(¹) This report was completed in February 2004. At that time the new Member States were still accession countries.

GSOEP	German socioeconomic panel
HPWS	High performance work systems
HRM	Human resource management
IAB	German Institute for Employment Research
IC	Intellectual capital
ICT	Information and communication technology
ISCED	International standard classification of education
IV	Instrumental-variable (method)
JCSEE	American Joint Committee on Standards for Educational Evaluation
LdV	Leonardo da Vinci
LFS	Labour force survey
LMP	Labour-market programme or policy
LTU	Long-term unemployment
MLM	Multi-level modelling (method)
NAPs	National action plans
NCDS	National Child Development Study
NEET	Not in education, employment or training (project)
Newskills	New job skill needs and the low skilled (project)
NIESR	National Institute of Economic and Social Research
NPM	New public management
OSLM/	Organisation for Strategic Labour
OSA	Market Research
PEZs	Prototype employment zones
PISA	Programme for international student assessment
PLMP	Passive labour-market policy

PPBS	Planning, programming and budgeting system
PURE	Public funding and private returns to education (project)
R&D	Research and development
ROCs	Regional training centres
RoR	Rates of return (method)
SEdHA	Socioeconomic determinants of healthy ageing (project)
SEM	Structural equation modelling (method)
SFB 186	<i>Sonderforschungsbereich 186</i> [Special research centre 186]
SME	Small and medium sized enterprise
TFP	Total factor productivity
TREE	Transitions from education to employment
TSER	Targeted socioeconomic research programme
VTLMT	Education, vocational training and labour-market transitions (project)
VTOS	Vocational training opportunities scheme
WEB	<i>Wet Educatie en Beroepsonderwijs</i> [Law on (adult) education and vocational education and training] (NL)
WVS	World value survey
YCS	Youth cohort study
Yuseder	Youth unemployment and social exclusion: dimensions, subjective experiences and institutional responses in six countries of the EU (project)
ZLSE	Zuercher longitudinal study

Contributions to the background report of the third research report

This synthesis report is largely based on the contributions to the background report published separately by Cedefop ⁽¹⁾. The contributions to the third research report have been regrouped into three broad themes, published in three volumes:

- (a) impact of education and training;
- (b) the foundations of evaluation and impact research;
- (c) evaluation of systems and programmes.

Impact of education and training

This volume addresses research work to assess the material and non-material impacts of education, training and skills at macroeconomic and regional level, and the impacts on enterprises and individuals. The picture is completed by case studies on the impact of active labour-market policies (ALMP) with training components, in terms of both effects for the economy and for social integration. It comprises the following contributions:

The impact of human capital on economic growth: a review	Rob A. Wilson, Geoff Briscoe
Empirical analysis of human capital development and economic growth in European regions	Hiro Izushi, Robert Huggins
Non-material benefits of education, training and skills at a macro level	Andy Green, John Preston, Lars-Erik Malmberg
Macroeconometric evaluation of active labour-market policy – a case study for Germany	Reinhard Hujer, Marco Caliendo, Christopher Zeiss
Active policies and measures: impact on integration and reintegration in the labour market and social life	Kenneth Walsh, H. David J. Parsons
The impact of human capital and human capital investments on company performance. Evidence from literature and European survey results	Bo Hansson, Ulf Johanson, Karl-Heinz Leitner
The benefits of education, training and skills from an individual life-course perspective with a particular focus on life-course and biographical research	Maren Heise, Wolfgang Meyer

(1) Descy and Tessaring, 2004a, b and c.

The foundations of evaluation and impact research

This volume gathers contributions on philosophical roots, types and standards of evaluation and impact research. It also discusses tools and methods for evaluating education and training systems and reforms. Competence measurement as well as designing and selecting key competences in an international context are also addressed. It comprises the following contributions:

Philosophies and types of evaluation research	Elliot Stern
Developing standards to evaluate vocational education and training programmes	Wolfgang Beywl, Sandra Speer
Methods and limitations of evaluation and impact research	Reinhard Hujer, Marco Caliendo, Dubravko Radic
From project to policy evaluation in vocational education and training - possible concepts and tools. Evidence from countries in transition.	Evelyn Viertel, Søren P. Nielsen, David L. Parkes, Søren Poulsen
Look, listen and learn: an international evaluation of adult learning	Beatriz Pont, Patrick Werquin
Measurement and evaluation of competence	Gerald A. Straka
An overarching conceptual framework for assessing key competences. Lessons from an interdisciplinary and policy-oriented approach	Dominique Simone Rychen

Evaluation of systems and programmes

This publication gathers contributions on recent evaluations of education and training systems and reforms. The case studies provide an insight into current evaluation practices in several European countries but also assess the results of international and EU supported programmes and initiatives. It comprises the following contributions:

Evaluating the impact of reforms of vocational education and training: examples of practice	Mike Coles
Evaluating systems reform in vocational education and training: learning from Danish and Dutch cases	Loek Nieuwenhuis, Hanne Shapiro
Evaluation of EU and international programmes and initiatives promoting mobility: selected case studies	Wolfgang Hellwig, Uwe Lauterbach, Hermann-Günter Hesse, Sabine Fabriz
Consultancy for free? Evaluation practice and culture in the European Union and central and eastern Europe. Findings from selected EU programmes	Bernd Baumgartl, Olga Strietska-Iliina, Gerhard Schaumberger
Quasi market reforms in employment and training services: first experiences and evaluation results	Ludo Struyven, Geert Steurs
Evaluation activities in the European Commission	Josep Molsosa

All these contributions can also be downloaded from:
http://www.trainingvillage.gr/etv/Projects_Networks/ResearchLab/

Preface

by Johan van Rens

In the 2000 European Council meeting in Lisbon and the following Councils of Stockholm, Barcelona, Bruges and Copenhagen, the Council set goals and objectives to facilitate progress towards a knowledge-based society and economy, through investment in human resources, Europe's main asset. The Education and Training 2010 work programme also places strong emphasis on more investment.

Cedefop, the reference centre on vocational training in the European Union, contributes to achieving European goals and objectives by facilitating the exchange of relevant knowledge and information on vocational education and training among stakeholders across Europe.

Publishing a report on research on evaluation and impact of education and training is an important aspect of that task. It contributes to increasing the understanding of the benefits of education and training – both material and non-material – for society, for companies and for individuals. It also assists discussion of what makes education and training and reintegration policies work.

Evaluation of education and training programmes and impact research are important strands of VET research, of particular relevance to policy-making. Evaluation contributes to the effective implementation of programmes and policies, ensuring that their objectives are met while achieving accountability criteria. Results of evaluation can feedback into policy formulation and programme design. Impact research, by investigating the benefits of education and training at different levels, provides a strong research background to support policy design and choices.

This report addresses the foundations, the methods, the results and the implications of evaluation and impact research. It will surely become a valuable reference for researchers and policy-makers in vocational education and training and can contribute to policy development via its findings and arguments.

Introduction

Aim of the reporting series on VET research

The present publication is the third issue in the series of reports on vocational education and training (VET) research that have been published by Cedefop since 1998 ⁽¹⁾. The aim of the reporting series is to provide a comprehensive overview of current research in initial and continuing VET in Europe, its results and their implications for policy, practice and future research. Attention is also paid to the theoretical and methodological foundations and due reference given to relations with institutional, economic, social, demographic and other fields of social action.

Publishing this reporting series, Cedefop intends to improve the transparency of VET research in Europe by fostering cooperation and communication both within the research community itself and between researchers, policy-makers and practitioners.

In contrast to previous editions, which attempted to cover the entire field of VET research, this third issue focuses on evaluation and impact research on education and training. This synthesis report has been elaborated on the basis of a number of contributions by researchers from various fields of education and training research and evaluation ⁽²⁾. We have pooled together these contributions and complemented them with additional research to review the various approaches, methods and results of evaluation and impact research. We also indicate the implications for policy and evaluation as well as discussing best practice evaluations.

In view of the wealth of evaluation and research material available, we cannot claim to have achieved extensive coverage of the subject but we aimed to present a representative sample, allowing the reader to gain an overview of:

- (a) the main functions and uses of evaluation, their philosophies and theories;
- (b) quantitative methods for evaluation and impact research at micro and macro level, as well as methods of qualitative inquiry and their application in evaluation;
- (c) tools for evaluating and comparing education and training systems, for accompanying the implementation – and assessing the outcomes – of programmes and reforms, illustrated by case studies;
- (d) impact research, quantifying the material and non-material benefits of education, training and skills at macro level of nations and regions, their contribution to enterprise performance and their influence on individuals' biographies and careers.

Our standpoint in compiling material for this report is that quantifying the contribution of education and training towards achieving a knowledge-based society and specifying its benefits at all levels – as in 'summative evaluations' and impact research – is as important as indicating ways to improve the design and implementation of education and training programmes or measures by 'formative evaluation'.

Report scope and definitions

Definition and role of VET and VET research

Broadly defined, VET comprises all more or less organised or structured activities that aim to provide people with knowledge, skills and competences necessary to perform a job or a set of jobs, whether or not they lead to a formal qualification. VET is independent of venue, age or other characteristics of participants and previous level of qualifications. VET may be job-specific or directed at a broader range of occupations. It may also include elements of general education. The major importance of

⁽¹⁾ Tessaring, 1998, 1999; Descy and Tessaring, 2001 a and b.

⁽²⁾ See Descy and Tessaring, 2004, three volumes, edited separately by Cedefop.

Table 1: **Selected terms related to evaluation and impact research** ^(a)

Term	Definition
Assessment	A general term embracing all methods used to appraise/judge performance of an individual, a group or a programme.
Benchmarking	A systematic process comparing the activities, processes and/or performance of a programme, organisation, country, etc., against a theoretical, political or existing reference with the aim of identifying ways to improve performance.
Effect	Socioeconomic change resulting directly or indirectly from an intervention.
Effectiveness	The extent to which the objectives of a policy or an intervention are achieved, usually without reference to costs.
Efficiency	Relationship between the results achieved (output) and the resources used (input).
<i>Ex ante</i> evaluation	An evaluation conducted before the implementation of an intervention to provide a prior assessment of whether socioeconomic issues have been diagnosed correctly, whether the strategy and objectives are relevant, whether it is coherent with other interventions or policies, etc.
<i>Ex post</i> evaluation	An evaluation conducted either on or after completion of an intervention or programme. It aims at accounting for the use of resources, effectiveness and efficiency of intervention and strives to understand the factors of success or failure.
Formative evaluation	An evaluation examining ways of improving and enhancing the implementation and management of interventions. Formative evaluations are primarily conducted for the benefit of those managing the intervention with the intention of improving their work or related future initiatives.
Impact	A general term used to describe the effects of a programme, policy or socioeconomic change. Impacts can be positive or negative as well as foreseen or unforeseen.
Input	The human, financial and physical resources used for an intervention.
Outcome	The positive or negative longer term socioeconomic changes or impacts that occur directly or indirectly as a result of an intervention's inputs, activities and outputs.
Output	Immediate and direct tangible result of an intervention.
Summative evaluation	A systematic investigation to determine the worth or merit of a programme, measure or policy using relevant social research methods and criteria, standards and indicators.

(a) These terms are extracted from the glossary in Annex 1.

VET for individuals, enterprises and society is widely acknowledged, and is perceived as a key element of lifelong learning.

Education and training policies, among others, have to consider the complex relationship between education and training and the socioeconomic system. It is the task of research to investigate and explain these relationships and their effects. It aims at reducing complexity and improving the understanding of causes and effects to identify the means and strategies expected to be most effective and acceptable in solving a problem. In that sense, VET research and evaluation

have common features because they are pragmatically oriented and meant to serve and inform policy- and decision-making.

Defining evaluation

As with most social science concepts and terms, the definition of evaluation is disputed. One of the main reasons is that evaluation has evolved through different historical periods and policy environments, with inputs from many disciplines and methodologies. The various perceptions of evaluation and its role are rooted in hard fought debates in the philosophy of science and theories of knowledge, related

to different value positions ⁽³⁾. In our publication, we will adopt the following definition, covering two dimensions of evaluation (Box 1 presents some other definitions put forward during the past 20 years in social research literature):

‘Evaluation is the systematic investigation to determine the significance, worth or merit of a programme, measure or policy by means of careful appraisal and study, based on relevant social research methods and criteria, standards and indicators (summative or impact evaluation).

Evaluation is at the same time a developmental process that illuminates or enlightens specific policies, processes and practice for its stakeholders, contributes to collective learning, reduces uncertainty in decision-making and helps to improve the design and implementation of the programme and/or future related initiatives (formative or process evaluation).’

Although they may sound synonymous, the numerous terms developed and used in evaluation and impact research reflect different aspects. More information will be given throughout this report when those specific terms are addressed. Table 1 provides a first overview of some key terms associated with evaluation research ⁽⁴⁾.

In this report, we mainly discuss evaluations undertaken for the sake of accountability, development, knowledge production or social improvement. However, it is worth mentioning another perception of evaluation, when the learning activity is not the object, but the subject of evaluation. Then we may discuss evaluation as a learning activity. Easton (1996: 15) has succinctly summed up this approach: ‘in short, evaluation is one of the prime mechanisms, if not the most important one, by which organisations and groups of people learn’. In Easton’s perspective, evaluation is about capacity building where the means become just as important as the end: the aim of evaluation activities is to generate in the organisation a spirit of constant enquiry that

can empower people to understand and hence do something about their own situation. This perception of evaluation as a tool for organisational learning is not discussed in this report ⁽⁵⁾.

Evaluation: is it research?

The relationship between evaluation and research is not clear-cut. Some argue that there is no distinction and that evaluation is a form of applied research; others insist in drawing differences.

Evaluation and research are forms of knowledge production and are similar in that they are based on theory and use scientific methods to discover, explain and develop new knowledge. Scientific evaluations comprise the full range of social-science research tools and apply the basic rules of collection and analysis of valid and relevant data in a systematic way. Good evaluation lives up to the same demands as research in terms of the representativity of the samples used, the validity and reliability of the findings and the transparency of the reporting. However, evaluation is a domain of applied research that is required to judge, value or assess concrete interventions or policies. In consequence, the evaluator can take over the criteria of the commissioner or target group, or define his/her own judgement criteria (Stockmann, 2000b: 11 ff.).

Smith and Glass (1987, cited in Cohen et al., 2000: 39) note that although there are many similarities between evaluation and research, there are also many differences. Among the most prominent of these, they note that evaluation, in contrast to research:

- (a) is carried out in order to provide a basis for decision-making in a concrete context;
- (b) is commissioned by a client and not initiated by the researcher;
- (c) is the property of the client, who may decide to make the findings known for a restricted audience only, whereas research is generally published and accessible to all;
- (d) operates within certain time constraints to

⁽³⁾ See Part 1 of this report and also for more details Stern, 2004.

⁽⁴⁾ See also the glossary in the Annex.

⁽⁵⁾ For more information on this conception of evaluation, the reader may refer to Easton (1996) and to Preskill and Torres (1999).

Box 1: Etymology and definitions of ‘evaluation’

Etymology of the term

A short analysis of the linguistic meaning of the term ‘evaluation’ in French, English and German semantics (French: *évaluation*, German: *Evaluation*, *Evaluierung*) may illustrate the origin of this term.

‘Evaluation’ is either rooted in lat. *valēre* – to be strong, to be worth, to rule – or *evallere* which means ‘to separate the chaff from the wheat’ (see Plinius 18, 10 (23) paragraph 97 and 99). Both terms come from Indo-Germanic **ual-* (to be strong, to govern) which is also the root for related terms in German, Slavic and other languages.

The Latin term, in a modified form and meaning, entered the French language in the 14th century with the verb *évaluer* (af: *avaluer*) which denotes the activity of valuing and estimating, mostly in an economic context. The term has two meanings: (1) to make a judgement on the value or price of something, for example a painting (*expertiser*); (2) to estimate a distance or quantity.

In medieval age, the term *evaluieren* (knowns: *Evaluation*, *Evaluierung*) entered the German language via the French *évaluer*, *valeur* with similar meanings (estimating, determining a value, calculating). The German terms *valor* (e.g. *Valorisierung*), *Evaluation* (outdated) and *Evaluation*, *-ierung* are assumed to have the same meaning. It is interesting to note that the ancient Old and Middle High German verbs *waltan*, *walten* equally belong to Indo-Germanic **ual-* (see above) and can be identified in Slavic and other languages as well. Germanic forms of *walten* are, for example, *verwalten* (to care properly, to keep order), *Valuta* (currency, standard, value) and *Gewalt* (force).

In the middle of the 19th century, the French term *évaluer* entered English literature as *evaluate* in two meanings: (1) in mathematics, to determine the ‘value’ of a quantitative term or to find a numeric term for quantitative facts or relations; (2) to calculate, determine or fix the total sum or final amount. The related known *evaluation* has two meanings: (a) the activity of estimating, valuing of products (a term related to this is *valuation*); and (b) the activity of calculating or determining the value of a mathematical term or physical quantity. Moreover, it can also denote the assessment of the significance of probabilities or proofs. (Kluge, 1999; Duden, 1989).

Some definitions of evaluation in the field social research

‘Evaluation measures the efficiency, effectiveness and responsiveness of a programme, carries out comparisons based on these measurements and uses the information achieved for finding political decisions and for the management of programmes. Evaluations should help managers to take decisions on programme prescriptions, rules

and technical support measures – at the same time they should support policy-makers in taking budget decisions and in legislative processes.’ (Wholey, 1984: 158, cited in Götz, 1999: 24).

‘Evaluation research focuses on the determination and explanation of the impacts produced by the implementation of programs.’ Evaluation research is generally defined as: ‘the systematic application of social research procedures in assessing the conceptualisation and design, implementation and utility of social intervention programs. In other words, evaluation research involves the use of social research methodologies to judge and improve the planning, monitoring, effectiveness, and efficiency of [...] human service programs.’ (Rossi, 1985: 19, cited in Stockmann, 1997: 99).

“Evaluation” refers to the process of determining the merit, worth or value of something, or the product of that process [...] The evaluation process normally involves some identification of relevant standards or merit, worth, or value; some investigation of the performance of evaluands on these standards; and some integration or synthesis of the results to achieve an overall evaluation or set of associated evaluations.’ (Scriven, 1991: 139, cited in Stern, 2004).

Similarly, Worthen et al. (1997: 5; cited in Hujer et al., 2004a: Section 2.1) defines evaluation as the ‘identification, clarification, and application of defensible criteria to determine an evaluation object’s value, quality, utility, effectiveness, or significance in relation to those criteria.’ Thus, evaluation ‘is the systematic investigation of the merit or worth of an object (program) for the purpose of reducing uncertainty in decision-making.’ (Mertens, 1998: 12, cited in Stockmann, 2000b: 13).

The most popular textbook definition of evaluation can be found in Rossi et al.’s book *Evaluation a systematic approach* (1999: 4, cited in Stern, 2004): ‘Program evaluation is the use of social research procedures to systematically investigate the effectiveness of social intervention programs. More specifically, evaluation researchers (evaluators) use social research methods to study, appraise, and help improve social programmes in all their important aspects, including the diagnosis of the social problems they address, their conceptualization and design, their implementation and administration, their outcomes, and their efficiency.’

The MEANS collection (EC, 1999, vol. 6: 17) defines evaluation as the ‘judgement on the value of a (public) intervention with reference to criteria and explicit standards (e.g. its relevance, its efficiency). [...] Certain definitions exclude the judgement dimension and limit evaluation to the measurement of the intervention’s effects. Other, more restrictive definitions, limit evaluation to the ex post estimation of effects.’

be ready for a specific decision-making process (timeliness) ⁽⁶⁾.

As Stern (2004) points out: ‘the boundaries between research and evaluation remain clouded. Many studies commissioned as evaluation contribute to knowledge production and are indistinguishable from research. Various distinctions have been proposed – including the short- rather than long-term nature of evaluation and its mainly instrumental intent. However, [...] many [...] studies [...] could be defined as research and, arguably, once a research-generated study is deployed for evaluative purposes, its character changes.’

Evaluation is therefore dual. This means that evaluation is part of empirical social research, whose theories and methods it applies. At the same time it is part of the political process that it influences with its results and can be used as an instrument of political steering. Thus evaluation may be oriented towards scientific standards and/or towards the information needs of commissioners or target groups. This duality requires the evaluator to find a balance between empirical and political objectives.

Evaluation and impact research

In this report, we have decided to compile results and discuss both evaluation and impact research, albeit separately. In fact, it is difficult to draw the line between the two. Often, ‘impact research or analysis’ is considered a type of evaluation, particularly associated with summative evaluation.

We follow a pragmatic approach and distinguish between:

- (a) ‘evaluation’ focusing on specific programmes and interventions, whether it happens before (*ex ante* evaluation); during (process or implementation evaluation) or after (*ex post* evaluation); conclusions from an evaluation are of direct use for policy-makers to decide to stop or to pursue a programme and by practitioners to improve it;
- (b) ‘impact research’ investigating the

relationship between education, training and skills, i.e. of human capital and various benefits for the individual, the enterprise or the economy or society more generally, irrespective of whether they are the result of specific programmes or interventions (e.g. the impact of the general rise of skills and qualifications on economic and social development).

Because of their different objects, evaluation and impact research use different approaches and methods. Evaluation broadly tries to estimate and to maximise the effect of a programme for participants using both qualitative and quantitative methods, mostly at microeconomic level. Impact research attempts to quantify the contribution of education, training and skills to economic and social development.

Finally, impact research does not strictly limit itself to VET but more generally investigates the benefits of education, training and skills. Unless specifically mentioned, it is this broader concept that we will use throughout the book.

Contents of the third research report

Evaluation is a systematic investigation to determine the significance, worth or benefits of a policy, programme or measure, using relevant social research methods, criteria, standards and indicators. At the same time, it is a developmental process that enlightens specific policies, processes and practices for its stakeholders. Ideally, it contributes to collective learning and to knowledge production. It reduces uncertainties in decision-making, helps to improve design and implementation of social interventions, while ensuring an effective allocation of resources. It contributes to more rational and reasoned policy-making and generates society debates.

It is therefore understandable that evaluation is spreading across Europe. It is used by

⁽⁶⁾ Following these differences, impact research (see below) is not evaluation, unless it is commissioned by a client and undertaken with a view to providing knowledge in a decision-making process. Cohen et al. (2000: 38) make this distinction between ‘evaluation’ and ‘evaluative research’ (i.e. impact research) even though they admit that there are many overlaps.

government and stakeholders as a tool for monitoring the implementation and maximising the allocation and use of resources in various social policies, including education and training and active labour-market policies. The European Commission also is evaluating its programmes and requires from the Member States follow-up and outcome evaluation for structural and social funds.

Beside the accountability or summative role attributed to evaluation, it is a tool for generating consensus among stakeholders and designing social interventions and reforms, in a constructivist or formative perspective. A characteristic of evaluation, compared with other fields of social research, is its direct links to policy- and decision-making.

The conclusions from impact research serve to understand better the relationship between skills and wealth (employment, productivity, economic growth) as well as well-being (reduction of crime, trust, social capital). Its results can rarely be used directly in policy but they contribute to informing it. In this report, we will concentrate on the impact of education, training and skills on material and non-material benefits at macro, individual and enterprise level.

To discuss these various facets in detail, present case studies and best practices, and review the methods used as well as the findings of evaluation and impact research, this research report is divided into four main parts.

Part 1 discusses evaluation itself. It reviews the phases and uses of evaluation according to the various stages of development of VET programmes: design, implementation and assessment. It argues that evaluations should include both formative and summative aspects depending on their adequacy for the various stages.

To understand better the various types of evaluations, it is essential to discuss the different basic philosophies underlying evaluation. According to positivist thinking, the most traditional school of evaluation, objective knowledge exists. It can be obtained through observation and causal relations can be deduced. Adherents to the positivist view of the world are more likely to carry out *ex post*

summative evaluation, trying to explain what works and what does not, i.e. to measure the outcomes, the benefits and impacts of policies and interventions. They tend to rely mostly on quantitative methods. ‘Scientific realists’ continue positivist thinking but are interested in opening the ‘black box’, i.e. understanding underlying mechanisms, processes and contexts which make programmes work or not. They tend to carry out *ex post* and implementation evaluation using both quantitative and qualitative methods of investigation to penetrate beneath the surface of observable inputs and outputs of a programme. ‘Action-theory’ evaluators are interested in the political rather than in the scientific use of evaluation. They take an interventionist role to help primarily all those participating in a programme or project to develop it according to their needs. Finally, ‘constructivists’ believe that reality is socially constructed. Constructivist evaluations involve all stakeholders in a development process, confronting their views to achieve a consensus, feedback to all, clarification of areas of agreement and disagreement in order to design interventions and policies.

To understand evaluation it is also useful to consider the level of evaluation activity in the various European countries and the factors that led to its development. Following Furubo and Sandahl (2002), European countries can be grouped in clusters according to their history and regular use of evaluation:

- (a) first wave countries (Germany, Sweden, the UK and the US) introduced and developed evaluation between 1960 and 1975;
- (b) second wave countries (Denmark, France, the Netherlands, Norway and Switzerland) developed evaluation as a governance practice in 1975 to 1990;
- (c) third wave countries introduced evaluation in from the 1990s onwards, for example due to external pressure from supranational organisations making it a condition for receiving financial support.

These clusters partly reflect the evaluative cultures as well as the maturity and recognition of evaluation activities across countries. It is national historical factors that determine the

degree of institutionalisation of evaluation as well as the political and administrative culture.

Evaluators are subject to pressures. First, time pressures because results need to be delivered quickly for the decision-making process. Second, political pressures because commissioners are often stakeholders of the very domain being evaluated and might be tempted to influence results. Third, as financial pressures may impose some limitations to the investigation, evaluation budgets are often only a small part of an intervention's budget. Despite these pressures, evaluators are expected to be as rigorous as social science researchers, while achieving results that are meaningful and can be of direct use for policy-makers. As a result, evaluators need to formulate standards to establish a reference point for their practice. Standards of evaluations thus serve as a framework for judging evaluations; as a protection that evaluators can evoke when being put under pressure; as a way of establishing a shared understanding between evaluators and clients as to what an evaluation is, and how it should be conducted; and as a tool for critical self-reflection. This discussion on standards concludes the first part of the report.

Part 2 discusses the objects, approaches and methods of evaluation and impact research. It is divided into two main chapters. The first chapter reviews the objects and the methods of programme evaluation. By 'programmes' we mean diverse measures and interventions of a limited duration, aiming to solve specific problems, and targeted at specific groups. 'Programmes' can be implemented at local, regional, national or international level. Typical examples are reintegration measures developed in the framework of active labour-market policies to combat unemployment and other forms of exclusion. Qualitative and quantitative methods can be used to investigate the implementation and results of these programmes.

Qualitative methods are used in programme evaluation to deepen understanding of the process and outcomes of a programme. Though they can be used on their own,

qualitative evaluation methods often usefully complement quantitative methods. While the latter are primarily used to generalise results and make them comparable across studies by using standardised variables and common dimensions, the former allow understanding of the mechanisms that lead to the observed programme outcomes. They also open the door to understanding the participants' experience in the programme. They capture the non-measurable effects.

Quantitative methods are used to measure outcomes and impacts of a specified variable, in this case training programmes for participants, and to compare them with those same outcome categories for people who did not participate. They provide information on accountability, efficiency and effectiveness of an intervention. The application of quantitative methods, however, encompasses several methodological difficulties that have to be overcome in order to measure the 'true' effect of an intervention:

- (a) distinguishing the programme effects from the effect of other factors;
- (b) determining what would be the hypothetical outcomes of the same people if they had not participated in the programme;
- (c) taking into account short-, medium-, and long-term outcomes;
- (d) avoiding selectivity and heterogeneity biases while ensuring validity.

To overcome these methodological difficulties, quantitative evaluation methods apply sophisticated evaluation designs and statistical models, which we review.

Qualitative and quantitative evaluation methods are not mutually exclusive but complementary. They offer different perspectives on the processes and outcomes of the programme under evaluation. In particular, their use is defined by the stages and aims of evaluation:

- (a) *ex ante* evaluations use mainly qualitative methods to determine whether the input conditions (human and financial resources, legal frame, theoretical basis, etc.) are suited to achieving the programme objectives;

- (b) both qualitative and quantitative evaluations can be used during implementation of programmes to investigate whether design, organisation and resources are appropriate, whether the target group has been reached, whether the objective set are appropriate, etc., and send feedback to programme managers;
- (c) after completion of the programme, primarily quantitative, but also qualitative, methods are used to investigate the effects and outcomes of the programme and conclude on its effectiveness and efficiency.

Chapter 2 of the second part is dedicated to the methods of impact research. For this report, we adopt a definition of impact research that is different from evaluation insofar as it does not focus on assessing the results of specific programmes but investigates the impact – both material and non-material – of education, training and skills on economic and social development as well as on company and individual performance. The chapter reviews the most common theories, objects, approaches and methods of impact research. It provides the theoretical and methodological background necessary to understand better the results of impact research that are presented in detail in Part 4 of the report.

The chapter begins by reviewing the theories and models used to investigate the determinants of economic growth and productivity. It discusses the theories that attempt to explain the mechanisms by which human capital affects growth: the neo-classical model and the new theories of endogenous growth. Beyond economic impact, however, there are other wider social benefits of human capital, such as reduction in crime, better health, trust, etc. Their nature and how they can be approached is discussed.

Skills development also has a major impact on company performance. However, quantifying that impact is difficult because of the numerous other intervening factors affecting productivity, profitability and competitiveness. Reviewed are the types of surveys used for collecting data in enterprises, the type of performance measures that are used to measure the outcomes of training and

the difficulties associated with developing a universal and relevant definition of training that can be applied across studies and enterprises.

The chapter on impact research ends by a review of research investigating the individual benefits of education and training:

- (a) research investigating the private returns on investment in human capital that assumes that the individual demand for education is based on an estimation of future gains or returns to that investment, either monetary (earnings) or non-monetary (e.g. avoidance of unemployment);
- (b) research using longitudinal data to measure the benefits to education and training in the life course, looking into long-term effects on transition or labour-market situation or comparing the situation of several cohorts to investigate the way social and economic changes affect the returns on education and training.

After these theoretical and methodological considerations, Part 3 discusses practically how evaluation can support the implementation and assessment of long-term education and training system reforms and of more targeted and short-term interventions such as active labour-market policies and programmes.

Before discussing evaluation practices, we considered it important to sketch the context in which education and training programmes and policies are implemented and the role they are expected to fulfil. The ageing of the European population, the increase in educational level of the population, the necessity of skills renewal through continuing vocational training (CVT) and adult learning, structural changes in labour-markets and their consequences in terms of skills mismatches and unemployment, all are among the pressures exerted on education and training systems. In response to these, reforms are implemented and new forms of learning are developed. Policy-makers can be supported by evaluators in this process:

- (a) in assessing the types of reforms and programmes needed, establishing priorities and choosing objectives;
- (b) in designing them while involving all stakeholders, deciding on types and levels of action and planning interventions;

- (c) to follow-up and steer implementation, ensuring that the original plan is followed, providing feedback, reorienting strategies of reform and ensuring a good allocation of resources;
- (d) to judge whether reforms and programmes have achieved their objectives and generated other unexpected or unintended effects;
- (e) to advise on future policies.

Education and training are complex systems that do not exist in isolation but have social and economic roles. Reforming education and training is a process that requires debate and compromise on what is to be changed and for what reasons, not only between those in the systems but also other stakeholders, especially social partners. It is essential to envisage and to evaluate the internal consistency of a new policy with other elements of the system (e.g. one cannot implement changes in curricula and in learning methods without involving – and properly training – teachers). But reforms have also to take into account the needs – and be coherent with the modes of functioning – of the production system and of society more generally (external consistency of a policy). It may sound trivial to say so, but the pedagogical logic of the education and training system must be articulated with the skill needs of the employment system.

Traditional evaluations of policies (or programmes) focus strictly on policy as a separate entity, treating it as if it was self-contained and independent from the historical, structural and institutional context. This reinforces the tendency to design policies that are limited in scope and do not take into account the existing institutions and interventions as well as the modes of functioning of economic and social systems. In this report, we favour a systemic approach, leading to more coherent approaches to education and training, and VET reforms.

Evaluation is a tool to serve policy-making in education and training. For example, for planning system reforms, the building blocks evaluations used by the European Training Foundation (ETF) provide a functionalist analysis of VET systems, taking into account different levels of action. They are used to

manage dialogue and reach consensus among stakeholders on the reforms needed. In addition, there are various tools that can be used to assess the results of a reform: questionnaires to participants and stakeholders, benchmarking and comparisons with other countries, external evaluations, macro-level evaluations of impact, etc. In doing this, it is particularly fruitful to constitute interdisciplinary teams of evaluators to grasp the complexity of system change. These considerations on the possible tools for evaluating education and training systems are complemented by a series of case studies and best practices. Discussed and analysed are:

- (a) the OECD policy review methods, where interdisciplinary teams of researchers undertake country visits, generating national and comparative evaluations as well as recommendations for future policy design;
- (b) the application of benchmarks in the EU, setting common targets for education and training policy in the framework of the Lisbon strategy, and in Denmark to identify areas for improvement and opening national debates;
- (c) the evaluations of VET systems reforms in Denmark and in the Netherlands, where the evaluation approaches were quite different; the former adopted a formative perspective, accompanying the process of reforms while the latter was an *ex-post* evaluation carried out four years after reform implementation;
- (d) the key role of teachers and trainers, illustrated by the example of Croatia, where they have been identified and involved as central elements in the process of reforming VET;
- (e) case studies of evaluations of PHARE VET programmes in the Czech Republic, Bulgaria and Slovakia;
- (f) evaluations of the Leonardo da Vinci programme, especially addressing mobility measures.

The last chapter of Part 3 discusses the role of training in combating unemployment and exclusion from the labour market, i.e. when training is used within active labour-market

policy (ALMP). The rise of unemployment and the tightening of public budgets generate a growing concern about the effectiveness of programmes designed to reintegrate people quickly into the labour market. Evaluations are therefore increasingly carried out to assess whether this objective has been reached and at what costs.

Training not being the unique form of ALMP, it must be considered whether it achieves the expected results in comparison to other types, such as job search assistance or subsidies to employment. The conclusion of the review of existing evaluations is that the effect of training on employment and earnings tends to be modest and that it is at the same time a very costly type of ALMP. In addition, it is not effective for all target groups, across regions or episodes of the business cycle. The conclusion might be that training does not work well enough and that other types of ALMP should be preferred. We argue that we need further evidence before drawing these kinds of conclusions. First, although they do not constitute a primary objective of ALMP, evaluations should be complemented by estimating the non-material benefits of training. Second, in order to design new and more efficient policies, policy-makers need to know ‘what worked?’ (or not) as well as ‘why?’ or ‘under what circumstances?’ Finally, also in the case of ALMP, a more systemic approach to evaluation should be envisaged because evaluating means considering whether a programme is successful, whether it can be replicated, articulated with other VET programmes, connected to employers’ hiring practices, or otherwise related to other established practices and institutions.

Finally, Part 4 discusses recent research results on the diverse benefits of education, training and skills for society, for companies and for individuals. Chapter 1 addresses the contribution of human capital and of social capital to economic growth at macro and regional level. The new theories of economic growth consider human capital investment, research and development (R&D) activities and the dissemination of knowledge as the main determinants of long-term economic growth.

OECD countries have invested heavily in education and training and this has had direct positive effect on labour productivity, i.e. on the growth rate of GDP per person employed. Generally speaking, a 1 % increase in school enrolment rates has led to an increase in GDP per capita of 1 % to 3 %. However, most studies provide only crude answers as to what type of skills, knowledge and competence and which level of education and training are the best ‘predictors’ of economic growth. It seems that the level of education will matter differently for countries with different levels of development: primary and (partly) secondary education contribute more in the poorest and intermediate development countries respectively and tertiary education influences GDP growth more in OECD countries. The quality of human capital (e.g. measured by school performance test) seems even more important than its ‘quantity’ (as measured by the years of schooling or level of education) in producing a positive impact on growth. Social capital and social infrastructure seem also to be important conditions that lead some countries or regions to accumulate more human capital than others and make them more efficient than others in using such inputs to enhance economic success. Finally, in addition to its direct positive effect on growth, human capital is also positively correlated with better health, crime reduction and social cohesion, making overall human capital a rather attractive investment. However, based on the results discussed in chapter two, it appears that raising education, skills and training levels is neither a necessary nor sufficient condition for generating macrosocial benefits. It is equally important to improve the distribution of educational outcomes, i.e. reduce the inequalities generated and carried over by education, to generate macrosocial benefits beyond the economic ones.

Increasingly firms are investing in all kinds of training, whether it is to develop skills that are only of use to the training firms or general skills that are more ‘marketable’. Results discussed in Chapter 3 demonstrate that investment in training does generate substantial gains for firms in terms of productivity, profitability and

stock market performance, independently of whether training is specific or general. Ensuring training outcomes, firm performance and innovation, seems to be influenced positively by human resource development practices and training need analysis. These results also apply to SMEs, although research on the links between training and economic performance is not so advanced in this case. Nevertheless, most investors and employers do not know about the payoff from training, leading to under-investment in it.

There are also considerable monetary and non-monetary benefits for the individuals to invest in education, training and skills (Chapter 4). As demonstrated by research on individual returns, there are positive impacts on earnings, labour-market participation and unemployment risk, which indirectly also generate other non-material benefits such as better health, parenting, crime reduction and social inclusion. Therefore, key policies are lowering the number of early school drop-outs and providing socially deprived groups with possibilities and incentives of continuing education and training. Life course and biographical studies demonstrate even further the relevance of education and training during the individual's life: initial participation in education and continuing training has cumulative effects for occupational career and

personal development. It is worth mentioning that the amplitude of this positive effect is increasing across generations. Nevertheless, selective access to CVT depending on previous educational level, gender, family background and other social factors contributes to the persistence and accumulation of social discrimination throughout the life course. This requires specific policies for fostering access by certain target groups to CVT.

In the Annexes to this report, the reader will find further information on various aspects of treated in the book, especially:

- (a) a glossary compiled by Cedefop of the main terms used in evaluation in English, French and German;
- (b) key European socioeconomic statistical indicators;
- (c) an example applied to quantitative evaluation methods;
- (d) a list of evaluation societies in Europe as well as of international evaluation societies and their contact addresses;
- (e) a list of the main evaluation journals in Europe and beyond.

This report also contains the extensive bibliography of sources consulted during its preparation, including references to all the contributions to the background report.

Understanding evaluation

Abstract

Evaluation is more complex than it appears at first sight. Various objectives and functions are assigned to it: improving decision-making by governments, employers, individuals and other stakeholders; improving the quality and efficiency of programmes; being a mechanism for public debate, social improvement and change; and serving accountability purposes.

Our perceptions of evaluation are influenced by our roles as stakeholders, participants in programmes, managers, policy-makers, financiers, evaluators or researchers. Our concepts of evaluation are also influenced by our disciplines and fields of specialisation as well as by the philosophy of science we adhere to and by our culture.

The first chapter describes the logic of vocational education and training (VET) programmes. It identifies evaluation as an accompanying tool in political processes and discusses how its various phases – design, implementation and impact – affect the use of concepts and methodologies. The chapter concludes by reviewing the functions attributed to evaluation.

Evaluation is embedded in philosophies and theories that shape how it is applied in particular contexts and the perceptions actors have of it and of its functions. This is the subject of Chapter 2.

Chapter 3 debates the historical development and intensity of evaluation activities as well as the diversity of evaluation cultures in Europe.

Chapter 4 discusses the standards applied to ensure that the conclusions and recommendations drawn from evaluations have been correctly reached, avoiding faulty or misleading findings, or compromises with those commissioning the evaluation.

Table of contents

1. Introduction to evaluation approaches and functions	21
1.1. Evaluation of VET programmes	22
1.2. Formative and summative: two evaluation approaches	23
1.3. Functions of evaluation	24
1.4. Pressures on evaluation practice	25
2. Philosophies and theories of evaluation	28
2.1. Philosophies of sciences and evaluation model	28
2.1.1. Positivism	28
2.1.2. Scientific realism	29
2.1.3. Action-theory	30
2.1.4. Constructivism	30
2.2. Two theories of evaluation	33
2.2.1. The programme theory	34
2.2.2. Theories of change	34
3. Evaluation traditions and cultures	36
3.1. Development of evaluation in various countries and organisational contexts	36
3.2. Intensity of evaluation activities	39
3.3. Evaluation culture in Europe	41
4. Standards for evaluation	42

List of tables, figures and boxes

Tables

1.1. Stage of VET programmes and their evaluation	23
1.2. Phase of political process and uses of evaluation	23
1.3. Typical functions of evaluation	25
1.4. Philosophies of science and characteristics of evaluation	32
1.5. Ranking of countries on the indicators of an evaluative culture	40

Figure

1.1. Functions of evaluations	26
-------------------------------	----

Boxes

1.1. Rules of evaluation according to the realist school	30
1.2. Hermeneutic cycle	31
1.3. New public management (NPM) characteristics	37
1.4. Standards and criteria	43

1. Introduction to evaluation approaches and functions

Although various forms of evaluation have been carried out since long ago ⁽¹⁾, its use in education and training started at the end of the 19th century when assessments of pupils or students were introduced (for a brief history, see Götz, 1999, Vol. 1: 27-29). In the 1920s, evaluators started to deal with pedagogical issues, such as the impact of learning materials, methods or curricula. These evaluations were used to decide on educational reforms and to back up political decisions on issues such as compensatory education, specific versus general education and individual versus group learning.

After the Second World War, 'modern' evaluation emerged. In the 1950s and 1960s, it was applied to development aid programmes to make costs more transparent. In the 1960s and 1970s evaluation extended to assessment and control of public policies and programmes. Among the first countries and governments which carried out such evaluations were the US – a pioneer in evaluation – Canada, Germany, Sweden and the UK, followed by most other countries in the 1970s and 1980s. The development of evaluation arose from the planning, programming and budgeting system (PPBS), introduced in the US in the 1960s and subsequently adopted by many other governments. The PPBS required the installation of effective mechanisms to control and prioritise public spending as well as to assist more rational and collective decision making. Later on, the introduction of the new public management (NPM) system, which entails decentralisation, tight public budgets and the evolution of the welfare State, required increased emphasis on programme reviews and evaluation, involving all stakeholders (Box 1.3, Section 3.1). These new forms of decentralised systems have not only increased the demand for evaluations of programmes and policies, but have also increasingly built evaluations into administration routines.

The current growth in evaluation activities results from both structural and managerial issues. In a context of malfunctions and polarisation of labour markets, in terms of career and earnings and of different access to education, jobs and technologies, the need for efficient measures to integrate excluded groups has increased. The increasing complexity of those structural socioeconomic difficulties requires multiple goal programmes and policies presenting complex organisational arrangements. At the same time, budgetary pressures have increased the need for improved performance, greater effectiveness and higher quality of public and private programmes and measures. International comparisons have also shed a new light on education and training systems and on the preconditions and processes that affect their performance. Finally, the distribution of the European Structural funds, including mandatory evaluation, has also contributed to the development of an evaluation culture in some countries.

Based on various kinds of data and taking into account benefits, alternatives and consequences, evaluation has become not only a management tool to ensure an efficient implementation of programmes and to measure their results but also an important instrument to generate empirical knowledge (Stockmann, 2000b: 11) and to transfer results into policy and practice. In fulfilling these functions, evaluation also raises questions and dilemmas (e.g. when benefits for one target group imply disadvantages for others).

In recent years, progress in research, coupled with the increased complexity of our societies, has led to diversification of evaluation issues. 'Once upon a time, the evaluator researcher read only the Bible ("Old Testament" [of evaluation], Campbell and Stanley, 1963; "New Testament", Cook and Campbell, 1979) to look up an appropriate research design and, hey presto, be out in the field. Nowadays, tyro investigators have to burrow their way through "sage" advice on "summative evaluation", "formative evaluation", "cost-free evaluation", "goal-free evaluation", "functional evaluation", "tailored evaluation", "comprehensive

(1) Assessments of soldiers' operational capability was, for instance, carried out in China in prehistory times.

evaluation”, “theory-driven evaluation”, “stakeholder-based evaluation”, “naturalistic evaluation”, “utilization-focused evaluation”, “preordinate evaluation”, “responsive evaluation”, and finally “meta-evaluation” before they even get their hands on a social program.’ (Pawson and Tilley, 1997: 1 f., cited in Furubo and Sandahl, 2002: 2).

1.1. Evaluation of VET programmes

A certain logic model underlies the conception of any VET programme, describing the connections between the programme input, activities and process (implementation), outputs, immediate outcomes and long-term impacts ⁽²⁾. The four stages of VET programmes proposed by Grubb and Ryan (1999: 13-19) ⁽³⁾ exemplify the logical model for VET interventions and programmes, see Table 1.1. At each of these stages, different forms of evaluation can be used, which relate to different evaluation approaches (see Section 1.2).

This model can also be looked at from a reverse perspective: if the desired outcomes (programme goals) include, for example, increased earnings and employment (stage 4), this supposes programme participants can develop their capacity to identify employment opportunities (stage 3), which requires developing the necessary skills by providing appropriate learning opportunities (stage 2), within a programme where human (teachers and instructors) and financial resources are accurately allocated (stage 1). The purpose of evaluation is to assess what went well and what went wrong in this fragile chain.

Stockmann (2000b) also identifies steps in the political process, to which different evaluation concepts correspond (Table 1.2):

(a) the phase of programme planning and preparation includes the conceptualisation, setting up of objectives and design of a planned intervention. Here, evaluation

has the task of investigating the material, personnel, institutional, financial and theoretical frameworks and input conditions of a programme. Such investigations are called *ex ante*, input or preformative evaluation;

- (b) during the implementation phase, i.e. during the execution of a programme, evaluation takes over control and advice functions. By collecting information on the programme development (process) and results, evaluation provides support for decisions which steer programme execution and enable in-time corrections to change the programme design. Such responsive evaluations that monitor the implementation and development of running programmes are called process or formative evaluation;
- (c) after the completion of the programme, evaluation has the task of recording and assessing the full extent of effects produced by the programme and uncovering interrelationships. Such evaluations are called *ex post*, summative or impact evaluations.

Thus, the objective of evaluation can be either oriented towards the improvement of programmes and of their execution or towards the analysis of their results. The cognitive interest in evaluation can be twofold (Jann, 1994: 311, cited in Stockmann, 2000b: 14):

- (a) improving policies and their implementation: ‘analysis for policy’ i.e. the improvement of future policies by application of scientific findings and methods (science for action);
- (b) analysing the results and impact: ‘analysis of policy’ whereby explanation and generalisation are the predominant objectives (science for knowledge).

⁽²⁾ See the Introduction for a discussion of these concepts.

⁽³⁾ Grubb and Ryan (1999) use the term ‘human capital development programmes’.

Table 1.1. **Stage of VET programmes and their evaluation**

Stage of VET programme	Objective and logic of action	Evaluation issues
Stage 1 Implementation	Formulation of the programme objectives and of its characteristics <ul style="list-style-type: none"> • resource allocation (administration, teachers, content, budget, participation, etc.); • legislative backup. 	<i>Ex ante</i> : achieve a consensus among stakeholders on programme objectives and characteristics. <i>Ex post</i> : determine whether a programme was established as intended and whether the allocation of resources was appropriate to the objectives assigned.
Stage 2 Learning process	Increase skills or competences of the participants <ul style="list-style-type: none"> • organise an appropriate teaching and learning process. 	Process: explore what kind of learning takes place, the appropriateness of the teaching methods, the relationship between learning and future outcomes; help the programme actors to achieve the objectives, etc.
Stage 3 Changing human behaviour	The skills acquired will change the behaviour of individuals on the labour market and in jobs.	Process/ <i>Ex post</i> : determine whether the programme has, e.g. laid the ground for more effective work or job-seeking by the unemployed or for an increase in productivity.
Stage 4 Creating long-term employment and non-employment outcomes	The changes in behaviour should result in economic and non-economic benefits for the individual, the enterprise or the organisation, the economy or/and the society.	<i>Ex post/impact</i> : <ul style="list-style-type: none"> • investigate the sustainability of programme effects, e.g. in terms of earnings increase and better job prospects; • take into account measurable outcomes as well as external effects and non-economic benefits (e.g. reduction of criminality, better health).

Source: Authors, based on Grubb and Ryan, 1999: 13-19.

Table 1.2. **Phase of political process and uses of evaluation**

Phases of the political process	Timing of analysis	Objective	Evaluation approaches
Formulation/planning phase	<i>Ex ante</i>	analysis for policy science for action	Preformative/formative: shaping the intervention
Implementation phase	On-going	analysis for policy analysis of policy	Formative or summative
Impact phase	<i>Ex post</i>	analysis of policy science for knowledge'	Summative: measuring outcomes; formative: providing feedback for follow-up projects

Source: Stockmann, 2000b: 15.

1.2. Formative and summative: two evaluation approaches

Formative evaluation collects evidence to redirect and improve interventions in the course of their execution (Plewis and Preston, 2001: 10). Therefore, it is mostly used in the planning, design and implementation phase of a

programme and uses mainly qualitative methods (see Part 2, Section 1.2). 'The organic and unpredictable nature of educational processes and outcomes suggests that interpretative methods such as ethnography, case studies, interviews and individual biography will be useful' (Plewis and Preston, 2001: 25). The results of formative evaluation allow for improving procedures, learning to do things

better, overcoming resistance, etc. (Stern, 2004).

Summative evaluation uses mostly quantitative methods to judge the outcomes, effects and impact of an intervention (see Part 2, Section 1.3). It is usually carried out *ex post*, after a programme has been completed, but can also be carried out during an intervention when first results are available. By providing respective feedback loops for the future of the intervention or for follow-up projects, summative evaluations also have a formative character. According to Plewis and Preston (2001: 10) summative evaluations highlight different policy options, e.g. whether the intervention should be continued or not, whether it should be replaced by something different or better, and whether its success merits extension to a wider population.

Scriven (1991) illustrates these two approaches as follows: ‘when the cook tastes the soup, that’s formative evaluation; when the guest tastes it, that’s summative evaluation.’

1.3. Functions of evaluation

Various functions, often overlapping or similar, are ascribed to evaluations in the literature reviewed. In this section we will review some of these before concluding on common functions usually attributed to evaluation.

According to Derlien and Rist (2002: 439 ff.) three basic functions – information, allocation and legitimisation – are ascribed to the evaluation of policies and programmes:

- (a) the information function aims to answer following questions: do policies work, what effects do they produce and are the programme objectives met? What unintended effects have occurred? How can policies be improved if evaluations have revealed shortcomings?
- (b) the allocation function aims to rationalise budget allocation: which programmes can be eliminated in case of negative evaluation results and what are the consequences? How can we get ‘value-for-money’ by reshaping individual programmes? This motive tends to emphasise quantity over quality;

- (c) the legitimisation function refers to the use of scientific evidence to justify policy decisions; however, it is always more difficult to prove the efficiency or effectiveness of programmes than to show the contrary, and thus much easier to delegitimise than legitimise policies.

Stern (2004) distinguishes four functions or purposes of evaluation:

- (a) accountability, i.e. providing evidence to sponsors and policy-makers of the achievements of a programme or policy, including its efficiency and quality;
- (b) development, i.e. improving the delivery or management of a programme;
- (c) knowledge production, i.e. developing new knowledge and understanding in the longer term. By investigating ‘what works’ and ‘what does not work’, by explaining causal links and validating knowledge;
- (d) social improvement, i.e. using empowerment and participative evaluation methods to improve the situation of the presumed beneficiaries of public interventions.

These four functions can be further elaborated in relation to programme stakeholders, their focus, main approaches and methods and key questions; see Table 1.3.

In the context of evaluating VET, Grubb and Ryan (1999) have also listed some basic functions of evaluation:

- (a) informing government decisions;
- (b) improving employer decisions about training;
- (c) informing individuals about their choices;
- (d) improving the quality of individual programmes;
- (e) triggering public debate about VET.

Grubb and Ryan (1999: 28) characterise this fifth function of evaluation to be ‘[...] less specific and purposeful, but perhaps more honest about what evaluation can accomplish.

The purposes we have already discussed are “rationalist” conceptions of evaluation: a programme is evaluated, positive or negative findings are then used to expand, modify, or shut down the programme, and evaluation results therefore have direct consequences. But practitioners of policy-oriented research have noted that their work is not often used in

this way [...]. Therefore a more modest conception of research and evaluation has developed, in which evaluation generates information that is then used in public debate, which may either act on evaluation results, reinterpret them, search for further information, or ignore them altogether. In this conception, evaluation is a way of discussing public programmes, where the information it provides is different from and possibly a corrective to the more political, self-interested, or value-laden positions that would otherwise have been taken.’

Yet another set of evaluation functions is proposed by Stockmann (2000b: 14 ff.) who ascribes evaluation four general functions (Figure 2.1):

(a) cognition: evaluations provide data that are directly relevant for policy decisions. Has

the measure reached the target group and met its needs? Has the programme been accepted? Are programme organisers capable of implementing the programme effectively and efficiently? Have the framework conditions changed? Which causal relationships exist? etc.;

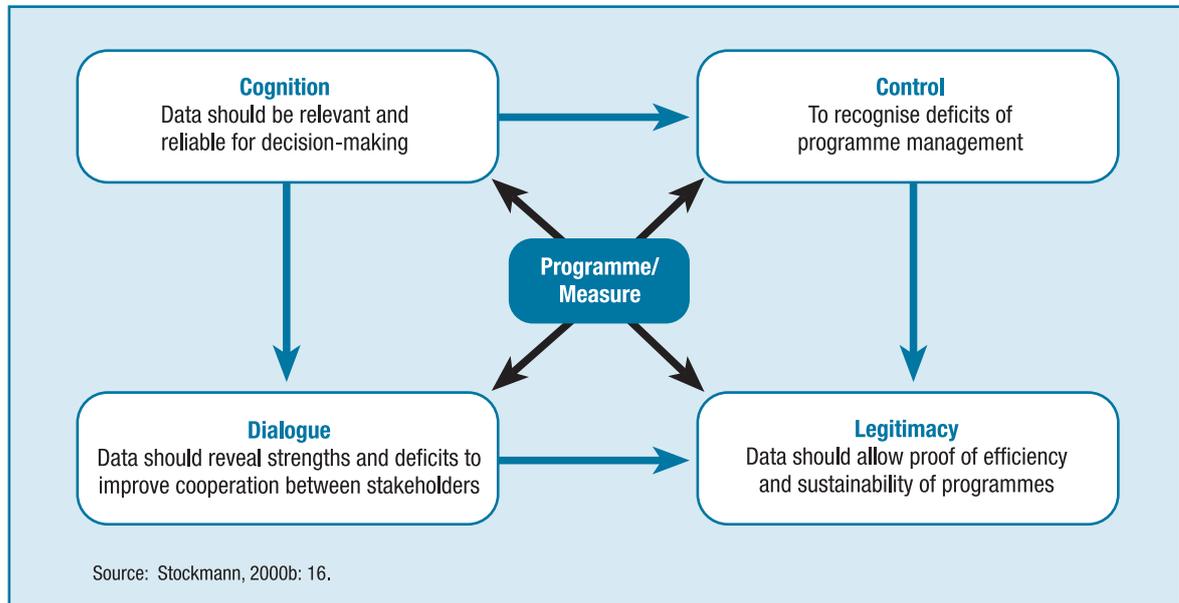
(b) control: evaluation is used to monitor a programme. The main interest is in recognising deficits in programme management and intervening quickly. At the same time, information is gathered to assess whether those involved in implementing the programme have performed their tasks;

(c) dialogue: evaluation supports dialogue between different stakeholders (donors, organisations in charge of execution, target groups and others affected) by providing

Table 1.3. **Typical functions of evaluation**

Purpose	Stakeholder	Focus	Main evaluation approaches	Key questions
Accountability for policy-makers	Parliaments, ministries, funders/sponsors. Management boards.	Impacts, outcomes, achievement of targets, value for money.	Indicators, performance measures, value for money studies, quantitative surveys.	What have been the results? Are they intended or unintended? Are resources well-used?
Development for programme improvement	Project coordinators, partner organisations, programme managers.	Identifying constraints. How they should be overcome? Delivery and implementation strategies.	Relating inputs to outputs, qualitative description, following processes over time.	How well is the programme being managed? Can it be implemented better?
Knowledge production and explanation	Programme planners, policy-makers, social partners. Academics.	Dissemination of good practice. What works and why? Organisational change.	Experimental, quasi experimental and non-experimental studies, case studies, systematic reviews and syntheses; meta-evaluations.	What lessons can be learned? Can these lessons be applied elsewhere? How should we do it next time?
Social improvement and social change	Programme beneficiaries and civil society.	To ensure full involvement, influence and control by citizens and affected groups.	Stakeholder involvement, participative reviews, advocacy.	What is the best way to involve affected groups? How can equal opportunities and social inclusion be ensured?

Source: Adapted from Stern, 2004.

Figure 1.1. **Functions of evaluations**

- information on evaluation results. This constitutes a transparent basis for judging success and deficits of the cooperation;
- (d) legitimacy: evaluation assesses which outputs and impacts have been achieved in relation to the inputs (personal and financial resources). In addition, *ex post* evaluations allow for a judgement on the sustainability of programme effects. On this basis, financiers and organisers can prove the efficiency of the financial means used and the effectiveness of the programme. Sometimes, however, evaluations also have a 'tactical' function and are used to legitimise certain political decisions. Evaluations are 'fashionable' and become decorative symbols for modern policy (Pollitt, 1998: 223).

From this short review, we conclude that evaluation has the following main functions:

- (a) improve decisions (by governments, employers, individuals and other stakeholders) by providing information on outcomes and performance in relation to the resources involved;
- (b) improve the quality and efficiency of programmes by investigating why a programme works or does not work with given inputs, and providing feedback;
- (c) be a mechanism for public debate and for

social improvement and change by being a corrective to the political or value-laden positions being adopted otherwise;

- (d) legitimate policy decisions.

Contrary to other authors, Dahler-Larsen (2004) argues that instead of describing functions that evaluation 'theoretically' has, one should consider those it actually has in concrete settings:

- (a) evaluation as control: evaluation is used by hierarchically superior authority to control whether a hierarchically inferior authority is carrying out its allotted tasks in an appropriate way;
- (b) evaluation as learning: when the programme managers or teachers, for example, can acquire, through the evaluation, new information and insight that can help them improve their practice;
- (c) evaluation used as strategy: evaluation is part of a hidden agenda and is used to legitimise policies and practices that have already *de facto* been decided on;
- (d) evaluation used as tactics: when evaluation is part of a dilatory tactic. Critical discussion about the object of evaluation is postponed until the evaluation has been carried out based on the argument that it must wait until the results are available;

- (e) evaluation as enlightenment: evaluation is used as a torch to illuminate and understand complex practices;
- (f) symbolic use of evaluation: when an organisation may initiate an evaluation procedure in order to give the impression to outsiders that it is prepared to look critically at its own practices, and to signal openness and a proactive attitude. The main aim of the evaluation here is to project a certain image of the organisation;
- (g) constitutive use of evaluation: evaluation is used to constitute the reality, which it is supposed to reflect. Through its description of an object, the object itself is defined.

1.4. Pressures on evaluation practice

Although evaluations, like any other kind of applied research, should follow the conventions for validation of social research, their functions in decision-making processes put pressure on evaluators. The main pressures come from:

- (a) the need for timeliness: evaluations, when used in decision-making, must deliver results in time, i.e. findings, conclusions and recommendations must be ready by the time a decision is to be taken;
- (b) the fact that evaluations are commissioned: evaluators are paid by a client who is often also a stakeholder in the domain evaluated, and who may try to influence the evaluator to bias the results.

Evaluators are therefore often under pressure. They have a deadline by which the findings must be presented, the resources available may impose certain limitations (e.g. on sample sizes, thus endangering representativity), and they may be unduly influenced by the client to use specific methods that give the whole project a certain bias, to repress certain findings, or to omit particular aspects of the evaluation (e.g. specific questions in a questionnaire survey) that might produce findings that are contrary to the interests of the client. In extreme cases,

'evaluations' are carried out when the result is already 'known', and where the exercise is merely some sort of charade to provide the necessary 'objective' pretext for an unpopular decision, or an attempt to justify a decision that has already been implemented.

Stronach and Morris (quoted in Cohen et al., 2000: 42) have argued that these pressures may result in evaluations becoming more 'conformative'. Conformative evaluations have the following characteristics:

- (a) they are short-term, taking project goals as given, and supporting their realisation;
- (b) they ignore the evaluation of longer-term learning outcomes, or anticipated economic/social consequences of the programme;
- (c) they give undue weight to the perceptions of programme managers who are responsible for developing and implementing the programme; as a result they tend to 'over-report' change;
- (d) they neglect and 'under-report' the views of classroom practitioners and programme critics;
- (e) they adopt a non-theoretical approach, and generally regard the aggregation of opinion as determining of overall significance;
- (f) they involve a close contractual relationship with the programme sponsors that either disbars public reporting, or encourages self-censorship to protect future funding prospects;
- (g) they undertake various forms of implicit advocacy for the programme in their reporting style;
- (h) they create and reinforce a professional schizophrenia in the research and evaluation community, whereby individuals come to hold divergent public and private opinions, or offer criticisms in general rather than in particular, or quietly develop 'academic' critiques which are at variance with their contractual evaluation activities, alternating between 'critical' and 'conformative' selves.

By relying on theories, using standards and applying methods of social science research, evaluators can avoid endangering the quality

and validity of evaluation. We will treat these questions throughout the report.

2. Philosophies and theories of evaluation

Because evaluation is embedded in general scientific research, it cannot be detached from debates taking place in the field. Therefore, it is important, first, to have an overview of various philosophical positions that result in different schools – or models – of evaluation, and, second, to understand some of the theories of evaluation that arise from these.

2.1. Philosophies of sciences and evaluation model

Positivism, i.e. the concept that objective knowledge exists and can be empirically verified, and constructivism, the idea that there is no ‘true’ knowledge and that all reality is socially constructed, constitute the two extreme philosophical schools on a continuum. Between these there are intermediary positions, such as scientific realism (which is, simply speaking, closer to positivist than constructivist thinking) and the action theory paradigm (closer to constructivism than positivism). Each philosophical position induces a different concept of research – and thus of evaluation – for those who adhere to it.

2.1.1. Positivism ⁽⁴⁾

Despite being criticised for some 50 years, the dominant school of evaluation is the logical positivism, which dates back to Comte (the first to use the term ‘positivistic’), Hume, Locke, Hobbes and Mill. Its core belief is that it is possible to obtain objective knowledge through observation, i.e. through empirical verification. ‘Evaluation information is considered to be “scientifically objective”. This objectivity is achieved by using “objective”

instruments like tests or questionnaires. Presumably, results produced with these instruments are reproducible. The data are analysed by quantitative techniques which are also “objective” in the sense that they can be verified by logical inspection regardless of who uses the techniques.’ (House, 1983: 51, cited in Stern, 2004).

For positivists, the world is a laboratory for social experiments and repeated observation of individual phenomena is the way to identify a regular pattern within a category of phenomena. ‘Internal validity’ is considered of crucial importance, i.e. establishing a relationship, and moreover a causal relationship, between the changes of a variable (e.g. an outcome measure such as earnings) and an antecedent variable (such as training) (Cook and Campbell, 1979; Scriven, 1991).

The term ‘positivism’ used by philosophers and social scientists derives from the acceptance of natural science as the paradigm of human knowledge. Giddens (1975, cited by Cohen and Manion, 1994: 12) identifies the following suppositions: ‘First, the methodological procedures of natural science may be directly applied to social sciences. Positivism here implies a particular stance concerning the social scientist as an observer of social reality. Second, the end-product of investigations by social scientists can be formulated in terms parallel to those of natural science. This means that their analyses must be expressed in “laws” or “law-like” generalisations of the same kind that have been established in relation to natural phenomena.’

Positivism is an important strand within the social and economic sciences and a working doctrine for the majority of economists as well as social and political scientists in North America and the UK. It is the basis for reductionism: the belief that it is possible to understand the whole by investigating its constituent parts. This view underpins many of the social experiments, surveys and economic models used in evaluation.

⁽⁴⁾ This section draws mainly on Stern (2004) and Stockmann (2000b: 17 ff.).

Opponents of positivism reject the belief that human behaviour is determined by general laws and characterised by underlying regularities. 'In rejecting the viewpoint of the detached, objective observer [...] anti-positivists would argue that individuals' behaviour can only be understood by the researcher sharing their frame of reference: understanding of individuals' interpretations of the world around them has to come from the inside, not the outside. Social science is thus seen as a subjective rather than an objective undertaking, as a means of dealing with the direct experience of people in specific contexts.' (Cohen and Manion, 1994: 26).

There is now widespread agreement that empirical work cannot rely only on observations collected by rigorous methods. It is impossible to observe empirically the entirety of any phenomena; all description is partial and incomplete with important unobservable elements. This is even more true for mechanisms which can not be observed. As Boyd (1991: 12) states, 'it is an important fact now, universally accepted, that many or all of the central methods of science are theory dependent.' This recognition of the theory dependence of all scientific inquiry underpins the now familiar critiques of logical positivism, even though there is considerable difference between the alternatives that the critics of positivism advocate.

2.1.2. Scientific realism ⁽⁵⁾

The scientific realism school believes in the possibility of accumulating reliable knowledge about the real world, albeit through different methodological approaches. According to Pawson and Tilley (1997) it seeks to open the 'black-box' within programmes or policies to uncover the mechanisms that bring about change. It does so by situating such mechanisms in contexts and attributing to contexts the key to what makes mechanisms work or not. This is especially important in domains such as VET where the evaluation 'objects' are varied and composed from

different elements assembled in configurations in various contexts (see Part 2, Section 1.1.1).

Rather than accept a logic that sees programmes and policies as simple chains of cause and effect, they are seen as embedded in multi-layered (or stratified) social and organisational processes. Evaluators need to focus on 'underlying mechanisms' i.e. those decisions or actions that lead to change, which is embedded in a broader social reality. However, these mechanisms are not uniform or consistent even within a single programme. Different mechanisms come into play in different contexts – which is why some programmes or policy instruments work in some but not all situations.

Like all those interested in causal inference, realists are also interested in making sense of patterns or regularities, not so much at the level of the effect of a programme but rather at the level of the underlying context in which mechanisms operate. Outcomes are the results of mechanisms unleashed by particular programmes in specific contexts. For Pawson and Tilley (1997), everything revolves around these 'context, mechanism and outcome' configurations. It is the mechanisms that bring about change, not all of which may be evident to programme architects or policy-makers.

According to the realist position, it is by examining and comparing the mechanisms and contexts in which they operate in relation to observed outcomes that it becomes possible to understand, as well as describe, success.

Policy-makers are then in a position to consider options such as:

- (a) narrowing down a programme to beneficiaries that are likely to change because of the mechanisms that work in the contexts they inhabit;
- (b) allowing for a greater differentiation and adaptability of a programme and its instruments to different contexts;
- (c) seeking to influence the contexts within which a programme seeks to be effective.

Box 1.1 provides a brief summary of the

⁽⁵⁾ This section draws mainly on the work of Pawson and Tilley (1997; cited in Stern, 2004) to describe the realist position in evaluation.

Box 1.1. Rules of evaluation according to the realist school

- Rule 1: **Generative causation**
Evaluators need to pay attention to how and why social programmes have the potential to cause change.
- Rule 2: **Ontological depth**
Evaluators need to penetrate beneath the surface of observable inputs and outputs of a programme.
- Rule 3: **Mechanisms**
Evaluators need to focus on how the mechanisms that usually generate social and behavioural problems are countered by the alternative mechanisms introduced by a social programme.
- Rule 4: **Contexts**
Evaluators need to understand the contexts within which problematic mechanisms are activated and in which programme mechanisms can be successful.
- Rule 5: **Outcomes**
Evaluators need to understand what are the outcomes of an initiative and how they are produced.
- Rule 6: **CMO configurations**
In order to develop transferable and cumulative lessons from research, evaluators need to orient their thinking to context-mechanism-outcome (CMO) pattern configurations.
- Rule 7: **Teacher learner processes**
In order to construct and test CMO pattern explanations, evaluators need to engage in a teacher-learner relationship with programme policy makers, practitioners and participants.
- Rule 8: **Open systems**
Evaluators need to acknowledge that programmes are implemented in a changing and permeable social world, and that programme effectiveness may thus be subverted or enhanced through the unanticipated intrusion of new contexts and new causal powers.

Source: Stern, 2004 (adapted from Pawson and Tilley, 1997).

realist position, in terms of eight ‘rules’ that are seen as encapsulating the key ideas of scientific realism and its method.

2.1.3. Action-theory ⁽⁶⁾

As a reaction to the methodological rigor implied in positivist and, partly also, in scientific realist thinking, a school of evaluators has focused on evaluations as a tool in political decision-making. They see evaluation as a political rather than a scientific means. Evaluations should, according to this view, primarily help all those participating in a programme or project. Evaluators should take an interventionist role and support all parties concerned with the evaluation.

The action theory – or participatory – paradigm extends this thinking and postulates that evaluations should not only be oriented towards quality control of innovations and interventions but at the same time allow ‘the construction, optimisation and legitimisation of model programmes’ (Lange, 1983: 256). This has several methodological consequences that are different from conventional social science research paradigms (Stockmann, 2000b: 18):

- (a) the primary objective is not falsifying theories or hypotheses but indicating alternative actions to solve existing problems;
- (b) the separation of evaluators and evaluation objects is avoided; evaluators abandon their detached position from the object under investigation and become partners of all those participating and concerned (evaluation becomes action research);
- (c) the research questions of the evaluator are not at the centre but rather the needs of the target groups;
- (d) the aim is not value neutrality; on the contrary, the evaluator’s role is to make values explicit;
- (e) the primary quality criteria of evaluation are no longer validity, reliability and objectivity, but communication, intervention, transparency and relevance.

2.1.4. Constructivism ⁽⁷⁾

Constructivists deny the existence of one ‘true’ reality and thus the possibility of objective knowledge about the world. Instead,

⁽⁶⁾ Cf. Stockmann, 2000b: 17f.

⁽⁷⁾ This section draws mainly on Stern (2004).

they assume that 'reality' is socially constructed by taking on different perspectives. In that they follow the tradition of certain continental European philosophers, in contrast to the mainly Anglo-Saxon school that underpins positivism.

The way we know, whatever the instruments and methods we use, is constructed by human actors or 'stakeholders'. According to Stufflebeam (2000: 71 f.), 'Constructivism rejects the existence of any ultimate reality and employs a subjectivist epistemology. It sees knowledge gained as one or more human constructions, uncertifiable, and constantly problematic and changing. It places the evaluators and program stakeholder at the centre of the inquiry process, employing all of them as the evaluation's "human instruments". The approach insists that evaluators be totally ethical in respecting and advocating for all the participants, especially the disenfranchised.'

In practical terms, the evaluator should be responsive to the 'claims, concerns and issues' of stakeholders and people affected: 'one of the major tasks for the evaluator is to conduct the evaluation in such a way that each group must confront and deal with the constructions of all others, a process we shall refer to as hermeneutic dialectic. [...] Ideally responsive evaluation seeks to reach consensus on all claims, concerns and issues [...]' (Guba and Lincoln, 1989: 41).

A distinctive role of the evaluator then is to help put together 'hermeneutic' circles (see Box 1.2) by bringing together divergent views and seeking to interpret and synthesise them to 'allow their mutual exploration by all parties' (Guba and Lincoln, 1989: 149). As Schwandt (1997: 69) has argued from a postmodernist standpoint, 'Only through situated use in discursive practices or language games do human actions acquire meaning'. Applied to evaluation he argues for the importance of the 'dialogic encounter' in which evaluators are 'becoming partners in an ethically informed, reasoned conversation about essentially contested concepts [...]' (ibid.: 79).

It should be noted that constructivist thinking is, to some extent, relevant to many contemporary evaluation challenges.

Box 1.2. Hermeneutic cycle

The hermeneutic approach covers a number of stipulations and various methods related to understanding and interpretation of phenomena (phenomenological approach). One important thinker in this field is Gadamer (1972). A hermeneutic approach seeks to elucidate what happens when one tries to understand a phenomenon in society. It can be used by evaluators of various philosophical schools, in particular action-theory and constructivism.

In a 'hermeneutic circle' one particular element is understood by the totality, while the totality is understood through its constituent parts. Interpretations achieve a more secure, detailed and varied understanding through a continued circular alternation between studying parts (system elements) and totalities (the total system).

An important point for discussion is whether objectivity is possible. Some advocates of the hermeneutic approach would argue so. Karl Popper, for example, claims that you can avoid wild interpretations and arbitrary postulations, which are a clear risk in hermeneutic understanding, by going through a number of well-conceived 'loops' for testing your interpretation hypotheses.

However, Gadamer questions this position. He insists that it is not possible to understand a phenomenon, for instance a foreign VET system, just by following certain methodical procedures. The interpreter cannot abstract him/herself from history and culture in trying to understand, as the researcher is her/himself part of history and culture. What the researcher/evaluator must do is to recognise and clearly articulate this tension.

Viertel et al. (2004) argue, that the phenomenological/hermeneutic method is primarily used for understanding rather than for explaining. Understanding and explanation are different ways of (re)cognition. Understanding is a more immediate experience or cognition of a phenomenon. If a phenomenon is difficult to understand, explanations are needed to comprehend fully, say, a component of a VET system and its precise role within the system. Explanations establish some distance from the phenomenon under observation. To explain something is to identify the causes behind the phenomenon under analysis, external to the phenomenon itself. To understand a phenomenon is to give a reason for it, which is internal. The relationship between understanding and explanation are tricky, one is not better or more correct than the other. Central in this reasoning is that understanding and explanations feed each other, that understanding is the point of departure for evaluations, and that explanations are needed when we fail to understand.

Source: Viertel et al., 2004, Section 3.1.

Table 1.4. **Philosophies of science and characteristics of evaluation**

	Positivism	Scientific realism	Action-theory	Constructivism
Nature of truth	A proposition that has not been tested empirically cannot be known to be true. Likewise, a proposition incapable of empirical test can never be confirmed to be true.	A proposition can be true in a specific context and not true in another one. Truth depends on specific configurations of context – mechanisms – outcome (CMO).	Truth is a common understanding of factual relations within a community of practice.	A proposition is neither tested nor untested. It can only be known to be true (credible) in relation to and in terms of informed and sophisticated social constructions.
Measurability and methodology	Everything exists in some measurable amount. If something cannot be measured it does not exist or cannot be considered. The inquiry seeks the verification of hypothesis, mainly using quantitative methods.	The inquiry seeks to open the 'black-box', i.e. to explain the CMO configuration beneath the surface of inputs and outputs. It uses quantitative as well as qualitative methods of inquiry.	Action research: the evaluator becomes an actor in the process under evaluation. The aim is to look for alternative actions, in collaboration with the actors of the domain, to solve existing problems. The needs of the actors, not the research questions of the evaluator, are in the centre. The inquiry relies on qualitative methods and measurability is not an issue.	Bringing together different views and seeking to interpret and synthesise them to allow their mutual exploration by all parties. Measurement may fit into some constructions but it is likely, at best, to play a supportive role. The evaluation relies on qualitative methods.
Evaluator posture and actions	'Disinterested scientist', being objective informer of decision-makers and changes agents. Action is not the responsibility of the evaluator, who views it as advocacy or subjectivity. Evaluator is distant from the phenomenon under observation in order not to jeopardise internal validity.	The evaluator may get involved with the target of evaluation with the aim of understanding the CMO configuration at stake. The evaluator is concerned with external validity. He adopts an objective stance while recognising the importance of context.	Evaluators abandon their 'disinterested' position and become subjective partners with all those participating. Political utility and participative evaluation. Evaluation is a political rather than a scientific act.	Evaluators are subjective partners, facilitators of the dialogue, and confrontation process amongst stakeholders. Evaluators seek not only dialogue but also actions from participants and stakeholders.
Independence of facts and values	Facts and values are independent of each other. Facts can be uncovered independently of the values that may later be brought to bear to interpret or give meaning to them. There are separate factual and valuational languages, the former describing '-isness' and the latter '-oughtness'.	Facts and values are related to each other. The task of the evaluator is to explain objectively how values and assumptions influence people's behaviour in particular contexts.	Facts and values are intertwined because the values of a community of practice will determine the kind of solutions that can be found to existing problems.	Facts and values are interdependent. Facts have no meaning except within some value framework; they are value-laden. There can be no separate observational and valuational languages.

Source: Authors based on Guba and Lincoln (1989: 103-111), Guba and Lincoln (2000: 163-188, cited in Nieuwenhuis and Shapiro, 2004).

According to Stern (2004) it is of particular relevance to VET because VET has many stakeholders. The entire system will only be able to progress if there is a broad consensus between them. For example, there may be a political desire to become more inclusive and involve previously marginal groups in training opportunities. The problem then is how to ensure that certain groups, such as women, young people and ethnic communities are given a higher profile in VET. Here, the involvement of many stakeholders will be inevitable. Furthermore the views of these stakeholders are more than data for the evaluator; they are the determinants and shapers (*Gestalter*) of possible action and change. Unless the trainers, employers, advocacy groups, funding authorities and employment services responsible for job-matching and the groups being 'targeted' cooperate, change will not occur. It is also likely that these stakeholders hold vital information and insights into the past experience of similar efforts, what went wrong and what did not, and what could be done to bring about improvements in the future ⁽⁸⁾.

According to Stern, the evaluator might then follow much of the constructivists' logic outlined above in carrying out the following tasks:

- (a) identify the different stakeholders who potentially have a stake in these areas of concern;
- (b) conduct a series of initial discussions to clarify what they know, what they want and what are their interests;
- (c) provide feedback to all stakeholders on their own and each others' interests, knowledge and concerns in a way that emphasises the similarities and differences;
- (d) clarify areas of agreement and disagreement and initiate discussions among the stakeholders and their representatives to clarify areas of consensus and continuing dissent;

- (e) agree what other sources of information could help move the stakeholders forward, perhaps by synthesising other available studies, perhaps by initiating new studies;
- (f) reach the best possible consensus about what should be done to improve VET provision and participation for the groups concerned.

Activities of constructivist evaluation emphasise the responsive, interactive, dialogic and 'orchestrating' role of the evaluator because stakeholders are considered a crucial source of data, sometimes more important than studies externally generating data. In this respect, although the 'battle of paradigms' in evaluation is not yet concluded and different views still exist, there is consensus that evaluations have to consider the perspectives and needs of stakeholders.

Guba and Lincoln (1989, 2000) have reviewed the main differences between the various philosophical schools of evaluation. Table 1.4, based on their work, concludes this section.

2.1. Two theories of evaluation ⁽⁹⁾

Recently there has been growing interest in the role of theory in evaluation. Hitherto, debates on evaluation focused on methods and techniques. To some extent, theories on evaluation reflect the debate between the different philosophies of evaluation sketched above.

Evaluation is applied research that fulfils a political function by answering questions such as: did the programme achieve what was intended; who benefited from it; could it be conducted better or more efficiently? All evaluations are based on data that can be interpreted in different ways. Theories provide a framework to support these interpretations. Further, evaluators will never be able to make a complete picture and theories can help fill in

⁽⁸⁾ See also Part 3 for further application in the VET domain.

⁽⁹⁾ This section is an extract from Stern (2004), chapter on evaluation theory.

the gaps in incomplete data. Similarly, theories can support prediction and explanation, which is their classic scientific role: to suggest and explain causal links and likely outcomes. Finally, many contemporary objects of evaluation are constructed. They are abstracted ideas which do not have a direct empirical referent. Learning organisation, an enterprise culture, the knowledge society, social capital, etc. are examples of constructed objects. Theories can make explicit the constructed objects of evaluation ⁽¹⁰⁾.

The following sections present two of the main theories underlying evaluation: the programme theory and the theory of change. In addition to these, every evaluation object or domain has its own theories. For example, theories of education and training are based on human capital theories and theories on human resource development, on learning and knowledge theories.

2.2.1. The programme theory

The programme theory, the dominant theory of evaluation, was developed by Scriven (1973, cited in Plewis and Preston, 2001: 20). He considers that evaluation should obtain more understanding of the effects of an intervention by opening up the programmes' 'black-box' and seeking to understand how programmes do or do not work. Chen (1990: 43) defines this as 'a specification of what must be done to achieve the desired goals, what other important impacts may also be anticipated and how these goals and impacts would be generated'.

This conceptualisation is linear and follows the 'treatment – implementation – outcome' logic: 'the causal processes underlying a program so that the reasons a program does or does not work can be understood' (ibid.: 191). The underlying logic of causality favoured by Chen is essentially consistent with classic experimental thinking. Following the principles of scientific realism discussed above, there continues to be an emphasis on the results of a programme or intervention

rather than on the context within which an interventions operates.

To overcome the limitations of programme theory, other theories are emerging on the issue of complexity in socioeconomic programmes (Sanderson, 2000). This is to take account of the fact that many interventions are not self-contained: they interact with other programmes and with other social and organisational processes. Thus, in VET, a new training system is embedded in an institutional and educational context which supports and constrains this system. Similarly, a training initiative at the level of a particular work group is mediated by the way work is organised, different management styles and labour market behaviour of employers and workers. Often there are multiple programmes and interventions operating simultaneously on a particular target group or area. The interaction of these various processes and programmes is one of the greatest challenges that evaluators face.

2.2.2. The theories of change

As early as in the mid-1970s, Weiss (1976) criticised simplistic notions of the instrumental use of evaluation, where it only serves to make recommendations that are then implemented. She argues for a complex understanding of decision and policy-making in which evaluation findings are internalised. This 'enlightenment view' of evaluation challenges the rational model of decision-making and is based on considerations of how organisations actually work. Weiss pleads for cumulative learning across many evaluations, by bringing together the various stakeholders.

This view underlies 'theories of change' which are the other main strand of theories on evaluation and are associated with the Aspen round table on comprehensive community initiatives. 'Evaluations that address the theoretical assumptions embedded in programs may have more influence on both policy and popular opinion.' They require programme practitioners to make their

⁽¹⁰⁾ For more information on these and other theories cf. Stern (2004).

assumptions explicit and to reach consensus with their colleagues and stakeholders about what they are trying to do and why (Weiss, 1995: 69).

By focusing on stakeholders and aspiring to encourage consensus among them, these evaluation theories share many features with the action-theory and even constructivist schools of evaluation. The 'theories of change' evaluation approach emphasises a collaborative, dialogic process. They take on board the themes of practitioners, make them coherent, explicit and testable and then seek to measure and describe programme outcomes in these terms.

Van der Knaap (1995) challenges the traditional rational-objectivist model of policy evaluation favouring rather a constructivist view in which policy-makers conduct dialogues about evaluation findings in order to reach their conclusions. Thus 'policy-making is conceived of as an ongoing dialogue, in which both governmental and societal actors contest their views on policy issues by exchanging arguments'. This argument challenges the 'positivist idea that policy evaluators may provide the policy-maker with neutral or objective feedback information or recommendations'. Rather than enlighten the policy-maker 'at best, the evaluator might contribute to the quality of policy discourse by entering into the negotiations that compose the policy-making processes with informed arguments and a willingness to listen, argue, and persuade or be persuaded'.

This shift from the rational to the argumentative is, according to Van der Knaap, a way to 'institutionalise policy orientated learning'. The evaluator is not relieved of the responsibility to provide reliable information and findings but needs also to supplement traditional analysis with material that will stimulate debate and allow different stakeholders to consider material presented from different perspectives. 'Rather than

regarding evaluative information as indisputable knowledge, it is viewed as a collection of arguments, which can be debated, accepted and disputed.' (Ville Valovirta, 2002).

In conclusion, theories of change, such as the systems approach ⁽¹¹⁾, see evaluation as a continuous and formative process of cumulative learning by all stakeholders, with evaluation results feeding back into the programme structure ⁽¹²⁾.

Preskill and Torres (1999) also attempt to revisit the use of evaluation in a comprehensive way. In their view, learning at an organisational level occurs as part of the evaluation process. This refers to constructivist theories of learning, in which learners engage in an active process of interpretation and dialogue in order to construct their own meanings. Preskill and Torres link this perspective with Lave and Wenger's (1991) notions of communities of practice, i.e. groups of individuals working together, who are interdependent and whose tacit knowledge and problem solving capacities are integrated into their social and professional life. They suggest that learning from evaluations 'will most likely occur when evaluation is collaborative, is grounded in constructivist and transformational learning theories, and builds communities of evaluation practice.' Stern (2004) notes that, if evaluation is to become part of an active learning process, this implies reinforcing the importance of developing communities of both evaluators and evaluation users. Communities of practice will interpret and construct their own meanings using data and findings and bring them into their own context.

⁽¹¹⁾ For a discussion of the various advantages and disadvantages of the systems approach cf. Plewis and Preston (2001: 23f.).

⁽¹²⁾ See also Part 3, Section 2.3 for the application of the systems approach to VET evaluation.

3. Evaluation traditions and cultures ⁽¹³⁾

3.1. Development of evaluation in various countries and organisational contexts

Evaluation did not develop at the same time across all countries, nor did it follow the same pathways with regard to fields of application or methodologies. International organisations such as the World Bank and the European Commission played – and still play – an important role in promoting evaluation. Several authors in the *International Atlas of Evaluation* (edited by Furubo et al., 2002) have traced the developments in a number of countries and organisational contexts ⁽¹⁴⁾.

The rise of evaluation as a distinct discipline can be explained by the emergence of factors influencing governing practice. These factors, however, did not appear at the same time in all countries, and their form has been moulded by the specific overall (administrative and) cultural context in which they appear.

In the US, evaluation started in the early 1960s with the War on Poverty initiative. Evaluations were based on positivist paradigms (see Section 2.1.1) and aimed primarily to link inputs to outputs, using quantitative methods and by being an integral part of the planning, programming and budgeting system (PPBS). Later, evaluations were increasingly based on applied social sciences and became a tool for policy-makers in assessing the continuing effectiveness and efficiency of programmes. In the 1980s, however, there was a systematic reduction in the use of evaluations and it is only recently that the federal government has revived them (Rist and Paliokas, 2002).

The evolution of evaluation usage in the UK during the 1960s and 1970s was triggered by similar factors. In the 1980s, programme analysis and review was abandoned and attention shifted to efficiency strategy and

financial management. Evaluation activity became fragmentary and closely linked to resource management. This situation changed in the late 1990s with the Labour government, which aimed to develop responsive and high quality services to reshape public expenditure and to launch cross-departmental initiatives. This led to increased evaluation of policy and programmes, modernisation of evaluation standards and tools and extension of evaluation into client participation (Gray and Jenkins, 2002).

Germany is one of the countries that institutionalised programme evaluation in the 1960s as a tool for managing the national government. Then, policy evaluation did not spread but remained concentrated in certain policy areas or sectors. Reunification in 1990 changed this situation dramatically. Evaluation became popular, particularly as regards the assessment of East German institutions. Increasingly tight public budgets also led to evaluations of practically all fields. Furthermore, evaluations were fostered by European regional funds – in particular for East Germany – and the new public management (NPM, see Box 1.3) (Derlien, 2002).

In Sweden, systematic evaluation developed in the mid-1960s although related activities have been used earlier. One important factor for the evolution of evaluation is the comprehensive ad hoc policy commission system where evaluations are an important part of governmental decision-making. Other factors are the highly decentralised public administration and the introduction of the NPM in the 1980s. (Furubo and Sandahl, 2002) The Institute for Labour Market Policy Evaluation created in the early 1990s currently plays a key role in fostering evaluation.

Evaluation in Italy emerged in the 1990s as a field of research and practice of government. The NPM and tight links to the European structural funds focusing on social cohesion training, job creation and territorial balance, made evaluation a prominent activity,

⁽¹³⁾ This chapter has partly been elaborated by S. Kristensen.

⁽¹⁴⁾ The following references refer to the contributions in this publication.

Box 1.3. New public management (NPM) characteristics

The reorganisation of traditional systems of administration and the introduction of economic steering methods stems from Anglo-Saxon countries. Its main objective is to implement market mechanisms in the process of public services. Harms (2000: 134f.) distinguishes between strategic and operational elements of NPM:

- strategic elements are decentralised responsibilities for resources; steering of outputs; and contract management;
- operational elements include introduction of commercial accountancy and cost service calculation; organisational development; reorientation of personnel management; and implementation of internal auditing, controlling and reporting.

A basic feature of the NPM system is decentralised resource responsibility, i.e. linking responsibility for production and financing to each other.

Buschor (2002: 61f.) summarises the main feature of NPM as follows:

- competition-oriented steering of public services, external as well as internal competition of schools; benchmarking and development of performance indicators;
- focus on effectiveness, efficiency and quality and performance;
- division of strategic management (e.g. government and parliament) and operative management;
- new equipment and budget: global budgeting where services and goals are defined and quantified;
- active promotion of innovation at each level of administration;
- promotion of delegation by decentralisation.

particularly for accountability and budgeting (Stame, 2002).

In Spain, interest in evaluation started in the 1980s because of a growing scarcity of public resources. This led to the establishment of evaluation institutions and observatories. Evaluations were focused on the measurement of objectives and performances, or on assessments of citizen's opinions. In the 1990s, there were increasing efforts to introduce programme evaluations, particularly prompted by ESF evaluations, decentralisation

and increasing demand for public services (Pazos and Zapico-Goñi, 2002).

Evaluation activities started in France in the 1960s with the PPBS, but were abandoned in the early 1980s. During the socialist governments in the 1980s a strong impetus was given to reform policies and related initiatives to measure policy effects. In the 1990s, evaluations developed further in the context of emerging counter-powers to the centralised national administration and executive. This led to reinforcing evaluation activities in the past 10 years. Evaluations are based on a pluralistic concepts and bodies, involving all relevant stakeholders. This development was also strongly influenced by European funds and the development of local autonomy (*déconcentration*) (Fontaine and Monnier, 2002).

In the Netherlands, the institutionalisation of evaluation started in the mid-1970s with attempts to link cutback management and budgetary review procedures. The 'new managerialism' gave evaluation the function of rationalising the reallocation of resources. This was accompanied by the development of basic sciences providing professional methodologies. Nowadays, evaluations play a major role in the policy process, because of NPM, quality management and audits (Bemelmans-Vidéc, 2002).

In the EU, the first steps to evaluation were taken in the early 1980s, starting with development aid, science and technology policy, and regional and social development through the structural funds. In the 1990s, most directorates general of the Commission carried out evaluations; some even had an evaluation unit. In particular the evaluation of RTD programmes ⁽¹⁵⁾ – based on regular evaluations and monitoring activities – can be seen as an antecedent of programme evaluation at European level, which is now an integral part of programme management. In 1995, a new evaluation scheme was adopted which combines a rolling system of annual monitoring exercises and periodic five-year

⁽¹⁵⁾ Community Research and Technological Development.

assessments of all research programmes as well as of the framework programme. It is a combination of *ex post* evaluation of previous programmes with an *ex ante* evaluation of future activities (Summa and Toulemonde, 2002).

The coming together of European labour markets, the increasing integration of European elements within labour-market policies and common challenges for education and training systems bring about new challenges and opportunities/requirements for evaluation. The Luxembourg process ⁽¹⁶⁾, for example, demands evidence of good practices, the development of common European indicator systems and evaluation standards relying on a broad spectrum of objectives, theoretical requirements and methodological approaches. Influences at national level, in terms of instruments used and their evaluation, are unavoidable (Brinkmann, 2000: 490).

The World Bank's evaluation activities date back to 1970 when an evaluation group within the World Bank's planning and budgeting system was established. The main principles were the close connection between evaluation and operational management, the systematic coverage of all lending operations, etc. Evaluations have evolved and grown with time and are now allocated at the Operations Evaluation Department (OED). Beyond project evaluation, the OED devotes a growing share of its resources to transversal studies, including process and impact evaluations, country policy assessments and global/regional policy reviews (Picciotto, 2002).

The diversity of evaluation history and use across countries is connected with the notion of government accountability, and is therefore closely associated with democratic methods of governance. In consequence, it appeared first in Western democratic countries, arriving

in central and eastern European countries behind the Iron Curtain significantly later. Furthermore, within Western democratic countries, Derlien (1990) distinguishes 'first wave countries' (Germany, Sweden, the UK and the US) that introduced and developed evaluation between 1960 and 1975; and 'second wave countries' (Denmark, France, the Netherlands, Norway and Switzerland), where evaluation was developed as a practice in governance between 1975 and 1990. Furubo et al. (2002), commenting on this categorisation, point to several factors – or predispositions – which can explain this sequence in the introduction of evaluation:

- (a) a political tradition that may be described in terms of social engineering, instrumentation and favouring rationalistic attitudes;
- (b) a tradition and belief in government intervention as a way of solving fundamental societal problems ⁽¹⁷⁾;
- (c) a high level of government expenditure;
- (d) the presence of civil servants with a background in the social sciences rather than law;
- (e) exposure to the US tradition of sociology ⁽¹⁸⁾.

Furubo et al., nevertheless, underline the complex nature of the mechanisms by which evaluations were introduced in Europe and the difficulty of making any rules in this respect. Sweden is somewhat atypical in comparison with the other first wave countries in being a small country with no federal structure. Denmark, on the other hand, shares many of the same characteristics as Sweden, but introduced evaluations over a decade later. The explanation of Albæk and Rieper (2002: 29) for this is that '[...] it seemed more sensible in the small unitary kingdom of Denmark to use alternative, less artificial and less time- and cost-consuming feedback mechanisms, such

⁽¹⁶⁾ In Luxembourg, in 1997, Member States agreed on the European Employment Strategy which is a form of 'management by objectives' whereby the Commission decides every year on employment guidelines. These guidelines are transformed into concrete policies in the Member States and reported on in national action plans. These plans are then assessed with a view to setting the next annual guidelines. For more details on this process and its results, see Section 4.3 in Part 3.

⁽¹⁷⁾ In the case of the UK and the US, it is a distrust towards all government interventions that renders their evaluation necessary.

⁽¹⁸⁾ The US displays an intensive funding of independent and scientifically founded evaluations of education and labour-market programmes. The main reason is that there is a deeply rooted mistrust in State control and regulation. The public expects to be informed on costs and returns on State policies.

as the close corporatist connections to affected interests'. It is interesting to note that this movement can also be reversible. In the case of France, Toulemonde (2000: 3) notes that the US-inspired evaluation model was done away with in the early 1980s 'however, the baby was thrown out with the bath water to such an extent that evaluation lost all significant support at government level for nearly ten years'.

Significant external pressures have also been exerted by donors and multinational organisations (such as the EU or the World Bank) to track and assess the outcomes and impacts of their projects, as well as to evaluate and report on performance. No country receiving assistance from these donors – in particular EU countries and candidate countries – can do so without an evaluation capacity.

As a result, a 'third wave' of evaluation in Europe appeared in the period from 1990 onwards. Whereas evaluation in the first and the second wave countries was a matter of internal predisposition, inspired by developments in the US, the third wave was caused by external factors imposing evaluations as an obligation. The decision of 1989 (reinforced in 1993) to make systematic evaluation a mandatory condition for receiving support from the Structural funds, has meant a quantum leap, both quantitatively and qualitatively, for evaluation activities in Europe. The EU has thus helped generate a market for evaluation and the creation of a 'community of practitioners' of evaluators in countries where evaluation would probably otherwise not feature as a part of regular administrative procedures. The European Commission has also contributed towards a certain convergence in Europe of practices and methodologies through its common guidelines for evaluations, notably the MEANS initiative (*Méthodes d'évaluation d'actions de nature structurelle*) (Molsosa, 2004).

After the fall of the Soviet Union and the centralistic forms of government, central and eastern European countries also started implementing evaluation activities on projects financed by external donors. The European Commission, the World Bank, the International

Labour Organisation (ILO), etc., have played a significant role in promoting evaluation in these countries (see Viertel et al., 2004; Baumgartl et al., 2004; Walsh and Parsons, 2004).

Health, education and development aid, plus social affairs and regional development, are the main areas for impact evaluation. 'These "soft policy areas" are [...], across countries, primary targets of evaluation and have, as such, influenced the development and direction of both evaluation theory and methodology. On a global scale, we can find no examples of where there is comparable influence of such sectors as defence, telecommunications, infrastructure, and energy on evaluation.' (Derlien and Rist, 2002: 452).

3.2. Intensity of evaluation activities

Evaluations in Europe are growing in number and slowly, but steadily, there is a greater supply of evaluators. National audit offices have been opened, while the attention attached to evaluations by parliament appears to be increasing. As Power (1998, cited in Leeuw, 2000: 75) has noted, we live more and more in an accountability society (or, as he calls it, an audit society). However, it must be noted that as interventions and programmes can produce unintended and undesired side-effects, this may also be true for evaluations. This makes it necessary that evaluators are involved in independent monitoring of their own activities (see Chapter 4 on standards and guidelines for evaluation.)

Based on country studies carried out by evaluation experts, Furubo and Sandahl (2002) use nine criteria to assess the maturity of countries concerning evaluation activities in general:

- I. evaluations take place in many policy domains;
- II. supply of domestic evaluators in different disciplines;
- III. national discourse concerning evaluation;
- IV. professional organisations concerned with evaluations;

Table 1.5. Ranking of countries on the indicators of an evaluative culture

	I	II	III	IV	V	VI	VII	VIII	IX	Total
US	2	2	2	2	2	2	2	2	2	18
CA	2	2	2	2	2	1	2	2	2	17
AU	2	2	2	2	1	1	2	2	2	16
S	2	2	2	1	2	1	2	2	2	16
NL	2	2	2	1	2	1	2	2	1	15
UK	2	2	2	2	1	1	2	1	2	15
D	2	2	1	2	1	1	2	1	1	13
DK	2	2	2	1	1	0	2	1	1	12
NO	2	1	1	1	2	1	1	2	1	12
F	2	1	1	2	2	1	1	1	0	11
FIN	2	1	1	1	1	1	1	1	1	10
CH	1	1	2	2	0	0	2	0	0	8
IRL	1	1	1	0	1	0	1	1	1	7
I	1	1	1	2	0	0	1	1	0	7
E	1	0	1	2	1	0	0	0	0	5
JP	1	0	0	1	1	0	0	0	0	3
Total	32	26	27	30	26	11	28	23	20	

2 high maturity or level of activity;
 1 medium maturity/level of activity;
 0 low or nonexistent level of activity.

Source: Furubo and Sandahl (2002: 10).

- V. institutional arrangements in the government for conducting evaluations and disseminating their results to decision-makers;
- VI. institutional arrangements in Parliament for conducting evaluations and disseminating their results to decision-makers;
- VII. pluralism of institutions or evaluators performing evaluations within each policy domain;
- VIII. evaluation within the Supreme Audit Institution;
- IX. proportion of outcome evaluations (concerning programme or policy outcomes) in relation to output and process evaluations.

After an interactive ranking process between the editors and the authors of each of the country studies, they weighted these criteria or indicators using scores (2 = high

maturity; 1= medium value, 0 = very low or non-existent level of activity). Table 1.5 (ibid.: 9) presents the result of this ranking.

Comparing Table 1.5 with another published earlier by Dierlen and Rist (1990), Furubo and Sandahl (2002: 11) conclude that there are a few countries in which the adoption of an evaluation culture took place in the 1960s or 1970s (Canada, Denmark, Germany, Sweden, the UK, the US). Only few more joined in the 1980s (Australia, the Netherlands). Then there is a sizeable group of countries which adopted an evaluation culture during the 1990s, or at least show some signs of having done so. In this category we find several European countries, such as Finland, Ireland, Italy, Norway and Spain ⁽¹⁹⁾.

The maturity of evaluation activities in different European countries is not only decided by the length of the evaluation tradition, but also by the scope of activities. As the example of France shows, evaluation

⁽¹⁹⁾ The EU does not make a separate entry on this list, but it is clear, as mentioned in the previous section, that the Commission plays a pivotal role in the development of an evaluation culture in Europe by making evaluation a condition for receiving support from e.g. the structural funds.

activities may be temporarily stopped, and the development slowed down, even though a country has been among the forerunners. Another factor to keep in mind is the possible uneven use of evaluation in different areas of government: 'Professional networks have remained highly compartmentalised and hardly inclined to bridge the gap with other sectors. In each political field experts have formalised their own evaluation techniques (e.g. expert panels for research, logical framework for development aid). Each field has generated its own network of experts and literature, without helping to generalise evaluation practice throughout the civil service' (Toulemonde, 2000: 4).

In his article on evaluation in Europe, Leeuw (2000: 60) uses the presence of national evaluation societies as an indicator of the scope of evaluation activities in a country ⁽²⁰⁾. The total number of evaluation societies in Europe is 15, in addition to the European Evaluation Society. However, the extent to which they actually can be said to signify the presence of a significant evaluation culture in a country would have to be judged not only on their existence, but also on the numbers and composition of their membership.

Concluding, countries of north and western Europe, the US and Canada are apparently the most active. This impression is confirmed in evaluation activities in education and training, regarding the scope of evaluation studies and material available for different countries ⁽²¹⁾.

3.3. Evaluation culture in Europe

The concept of 'culture' is notoriously difficult to come to terms with. Applying it to evaluation in a European context does not make it any easier.

Based on case studies of evaluations carried out mainly in central and eastern European countries, Baumgartl et al. (2004) note that 'it is clear from our cases that the evaluation culture varies strongly across countries. Sometimes evaluation is perceived as a threat, though at others it is welcomed as a form of free consulting'.

Toulemonde (2000: 10), in a paper on evaluation cultures in Europe, states that 'the diversity of our European nations is creating and will continue to create a wide variety of administrative settings. A pragmatic fit to these multiple realities is likely to generate multiple forms of evaluation. This will create numerous opportunities to innovate in the design, performance or use of evaluation'. Toulemonde also refers to the cultures of individual evaluators across Europe: 'To date, differences of culture between consultants in the same country are huge, and no correlation has appeared between the quality of evaluation work and the nationality of the evaluators. On the contrary, the authors of the best reports seem to share the same professional qualities, whether they are Spanish or British' (ibid.: 7).

Stern (2004: Section 8) notes that evaluation cultures in Europe are not yet converging: 'Alongside the cultural diversity of Europe, and not disconnected from it, is the cultural diversity among evaluators. Within the social and economic sciences in particular, there are familiar cleavages [...]. For example, evaluation also has its advocates for largely empirical and positivist methodologies as well as advocates for theory-based investigation in various forms. What constitutes evidence, validity, generalisability and legitimate conclusions are hotly debated among evaluators as they are in other academic and professional circles. To an extent, these cultural divides among evaluators mirror and are reinforced by cultural divisions within Europe. The philosophical patrimony of Latin,

⁽²⁰⁾ A similar indicator is also used by Furubo and Sandahl (criteria IV).

Annex 5 presents a list and contact address of these evaluations societies, by country. Also listed are international evaluation societies such as the European, the African and the American evaluation societies.

⁽²¹⁾ Almost all experts contributing to the background report of this research report complained of the lack of studies in southern European countries, except studies which evaluate ESF programmes and are mostly commissioned or required by the donor (European Commission).

Scandinavian and Anglo-Saxon countries, can make them receptive to different evaluation approaches and methodologies. There is also a history of policy borrowing within Europe, with established patterns of shared professional and policy networks that predispose some countries to adopt practices more easily from particular countries rather than others (Italians are more likely to “borrow” from the French, Scandinavians from each other and the British from North Americans). So, the alternatives of convergence or national – or at least regional – specificity are available to evaluators as they are to Europe’s citizens and their Member States. Among the influences that will shape these alternatives, the growth of evaluation societies at a European level and the policies of the EU, are likely to be important as well as the dissemination of evaluation standards and procedures by European institutions.’

Finally, a white paper of the European Commission (Schmitt von Sydow, 2002:9) ⁽²²⁾ mentions that between 1997 and 2000, in the Commission, an evaluation culture favouring mid-term formative evaluations existed. Only a few evaluations are summative as yet and *ex ante* evaluations are also not a common practice. At the same time, the integration of stakeholders is rather weak in European Commission evaluations. The white paper states that it is still too early to speak about a general evaluation culture. The most common purposes of evaluations are ‘to enhance democratic accountability, to assist political decisions about legislation, policies and programmes, to promote closer understanding between stakeholders and to support the implementation and management of existing programmes’ (Schmitt von Sydow, 2002: 20 f.). General rules or standards are considered to be more effective for ensuring objectivity and neutrality than, for example, the establishment of a formal independent evaluation function within the Commission (cf. also Molsosa, 2004).

In conclusion, European evaluation culture is characterised by diversity, which may be seen either as a weakness or a strength. It is clear, however, that there are countries in Europe where evaluations are more widespread and accepted than others, where evaluations happen as a regular feature of administrative practices, and where the community of practice of evaluators is broader. In some countries, courses for evaluators are offered at university level and special chairs at universities are devoted to evaluation; this factor could also be considered as an indicator for the presence of an evaluation culture. Historical factors and the overall political and administrative culture determine the degree of institutionalisation of evaluation at the national level.

4. Standards for evaluation ⁽²³⁾

Since evaluations are used to validate practical decisions, one must ask how they are validated; i.e. how it is ensured that the conclusions and recommendations arising from them have been reached correctly. Faulty or misleading evaluation findings, when used in a decision-making process, may have serious consequences for people involved in, or affected by, the programmes evaluated. Rossi et al. (1999, quoted in Stern, 2004) evoke the close relationship between evaluation and research, and tackle the question of validation by defining evaluation as ‘the use of social research procedures to systematically investigate the effectiveness of social intervention programmes’. Evaluations, according to this understanding, are validated and should be judged according to the same standards as any other form of social research.

The pitfalls of evaluation activities are often a direct consequence of the evaluation functions when projects are situated in concrete, politicised contexts (cf. Dahler-Larsen, 2004; Section 1.3). The formulation of

⁽²²⁾ This white paper has been elaborated, based on four hearings, by 27 staff members of the European Commission (members of the Working Group 2b) and 18 external evaluation experts from a number of European countries.

⁽²³⁾ This section has been drafted mainly by S. Kristensen, drawing on Beywl and Speer (2004) and Stern (2004) among others.

standards for evaluations can be seen as a response to these drawbacks. The very fact that there is a perceived need to formulate specific standards for evaluations is a sign that the practice of evaluation has a problematic relationship with the conventions of validation of social research.

Standards of evaluations thus serve as:

- (a) a framework for judging evaluations (conducting meta-evaluations);
- (b) a protection that evaluators can evoke when being put under pressure to comply with the views of certain stakeholders;
- (c) a way of establishing a shared understanding between evaluators and clients as to what an evaluation is, and how it should be conducted;
- (d) a tool for critical self-reflection.

Box 1.4. Standards and criteria

As Beywl and Speer (2004, with reference to New Webster's Dictionary, 1989) note, the word 'standard' has a number of different meanings. In general, it means norm, rule or yardstick and implies some kind of comparison, e.g. behaviour or a product, programme, etc., that should be set against or follow a normative [or ethical] rule. In evaluation terminology, 'a standard is an authoritative principle or rule that usually implies a model or pattern for guidance, by comparison with which the quantity, excellence, correctness, etc. of other things may be determined. [...] A criterion is a rule or principle used to judge the value, suitability, probability, etc., of something without necessarily implying any comparison.'

Because evaluators 'judge' a programme, controlling and questioning their roles is unavoidable. As with any other debate about the responsibility of professionals, the question of *quis custodiet ipso custodes* arises. The subsequent question is, through what means and against what criteria and standards does regulation occur? To establish some shared agreement among evaluators, commissioners and those affected by evaluation, evaluation codes and standards

have become prevalent. However, codes and standards also have a broader purpose. In a decentralised system composed of many stakeholders, standards are a way of regulating behaviour across organisational boundaries, provided that all parties accept these norms (Stern, 2004). Standards for evaluations and guidelines for evaluators formulate organisational, legal, technical or methodological requirements, but also ethical principles and considerations (Beywl and Speer, 2004).

Various attempts have been made to formulate standards for evaluations (and evaluators), one of the earliest and most influential being led by the American joint committee on standards for educational evaluation (JCSEE) in 1994. The standards were originally developed for use in an educational context, but have since been used in a broader context to cover all fields where evaluations are undertaken. In Europe, the German, Swiss and Austrian evaluation societies have used these as the basis to develop their own national evaluation standards. The JCSEE standards operate with four different 'clusters' of programme evaluation standards that are concerned with the utility, feasibility, propriety and accuracy of evaluations respectively ⁽²⁴⁾.

These four clusters of standards are concerned with the following aims:

- (a) utility standards; the evaluation should deliver the information that the client needs, and at the stipulated time and in an appropriate form;
- (b) feasibility standards; the scope of the evaluation should have a realistic relationship to factual needs and economic aspects, and be conducted in a diplomatic and prudent way;
- (c) propriety standards; the evaluation should respect the personal rights of the people who are involved in the evaluation as well as those affected by the results;
- (d) accuracy standards; the evaluation should

⁽²⁴⁾ Each cluster contains a number of standards that relate to each aspect. For example, in the German standards, which are modelled closely on the American standards, the four clusters of standards comprise 25 individual standards. See for details the contribution by Beywl and Speer to the background report (2004).

produce valid and adequate findings by means of methodologically correct procedures.

Standards do not prescribe particular methods, but stipulate that in any phase of the evaluation relevant and appropriate methods should be employed. Standards, therefore, do not subscribe to any particular epistemological branch of social research, but are equally applicable to evaluations informed by, for example, positivist and constructivist traditions.

In connection with the development and use of standards, there are three questions that should interest us particularly:

- (a) are these standards relevant also to evaluation of VET?
- (b) to what extent are they transferable across cultures?
- (c) are they actually used by evaluators and commissioners?

As for the first question, Beywl and Speer has conducted a survey among VET-experts (in Germany and Austria) to assess the relevance of the standards for evaluations of VET-projects and programmes. His conclusion is that there are no difficulties in using these standards here. Since the original American standards were developed for use in an educational context, this is perhaps not surprising.

The transferability of standards across cultures is a more problematic issue. Because of different institutional settings and differently developing markets for evaluation, some evaluation standards may be culturally sensitive. According to Beywl and Speer (2004, Section 3.1), evaluation activities in the US, and particularly the evaluation standards developed, have undoubtedly influenced European evaluation initiatives. This can be interpreted as a general dominance of American evaluation. Beywl and Speer (2004: Section 3.1) notes nevertheless that 'professional standards are usually shaped by values and norms, which can vary widely from culture to culture. In addition, the configuration of the parliamentary system is an important determinant of national evaluation culture'.

Even though the American standards have been adapted to meet the national standards in Germany, Austria and Switzerland, other countries in Europe have chosen to develop their own. This is the case in France and the UK (still at the development stage) as well as in Italy and Finland. Also, the European Commission and the International Labour Organisation (ILO) have developed their own norms and guidelines for evaluating programmes and projects that have received funding from them. Individual evaluators may also come up with their own standards (see for example Coles, 2004; Baumgartl et al., 2004).

Beywl and Speer (2004: Section 3.2) concludes that there are no serious discrepancies or contradiction between the German standards (modelled on the American ones) and the other European standards developed independently. Seen in a world-wide perspective, however, it is only in America, Europe and Australia where evaluation standards have been developed.

Cultural differences may not only find an expression in the formulation of standards. Significant differences may also become evident in the weight that is placed on individual standards, or clusters of standards. This is an element that has not been treated in any of the contributions to the background report (Descy and Tessaring, 2004 a, b and c), and we are not in a position to judge whether these differences in priorities are phenomena that should be ascribed to differences in national (evaluation) cultures, or whether they merely reflect the different dispositions of individual researchers or the circumstances under which they write. Is it significant, then, when Stern (2004) tends to consider utility standards as the most immediately relevant, Hellwig et al. (2004: Section 1.3.2) in their contribution state that 'with regard to evaluation quality, central to this project, accuracy standards will take priority, as the essential criteria for evaluation appraisal may be deduced from them.'?

As for the final question about the actual use of evaluation standards, Stern (2004: Section 7.1) notes that 'while there has been

a great deal of discussion around standards and codes in North America, Australasia and, most recently, in Europe, the evidence of take-up of such standards and codes is sparse. Most evaluators can come up with examples of “breaches” in most of their evaluations. The systematic application of standards is rare’. This may be partly explained by the fact that the standards of evaluation are formulated by national evaluation societies which, in many cases, only comprise a fraction of all the stakeholders in the field and mainly the evaluators themselves, not the commission-

ers. Second, built into the standards is what Beywl and Speer (2004: Section 2.2) termed an ‘unresolvable conflict’ between the strict adherence to methodologies of social research and the requirements of those who have commissioned the evaluation. Adhering rigorously to all four clusters of standards is often not possible, and it is up to the evaluator to strike an appropriate balance between the intrinsic demands of social research methods and the extrinsic demands coming from the environment and the users of the evaluation.

Approaches and methods of evaluation and impact research

Abstract

There is general agreement that the outcomes of social programmes should be judged by evaluation. Commissioners of programmes increasingly ask for evidence of efficient use of the means invested and enquire whether the programmes were successful or not. Evaluation is, however, a complex task and requires several steps. The range of topics for investigation is virtually unlimited and, since every topic requires a specific methodological approach, there exists no general evaluation strategy. To be successful, every evaluation must pre-specify a set of preliminary aspects. It must be stated precisely which object should be evaluated, what are the intended effects and outcomes and when – or at what stage – the evaluation should take place. Furthermore, the evaluation criteria and the most suited procedure and methods for evaluation must be determined. The design of evaluation and the methods and approaches chosen should be appropriate to each stage of a programme: planning, implementation and completion, as well as suited to the pre-specified objectives of the programme.

Impact research investigates the impact of a particular phenomenon – for example educational expansion, demographic development, technology diffusion, structural change – on individuals, companies, labour markets, economies and social systems. Because most of these phenomena are not planned and executed by a particular agent (as with, for example, an educational programme) but represent elements of the broader notion of social change, their impacts have in principle no limited duration and are not restricted to specific target groups, regions or periods of time. Impact research constitutes a broader field than evaluation research; it has its own problematic and has developed its own approaches and methods. Addressed in this part are the methods to research the impact of education, training or skills on economic growth and wider social benefits of, and the material and non-material returns on, education and training investments for individuals and companies.

This part includes two main chapters: approaches and methodologies of programme evaluation and approaches and methods of impact research.

Chapter 1 starts by delimiting the various types of objects, effects and outcomes of a programme under evaluation. Then, the main part of this chapter is devoted to methods that are usually applied when designing, implementing and evaluating the effects and outcomes of VET programmes. First, methods of qualitative inquiry and their application in evaluation are discussed. Second, quantitative approaches are presented, discussing evaluation challenges and possible biases as well as methods that are used at micro and aggregate levels.

Chapter 2 gives an overview on the approaches and methods used to research the impact or benefits of education, training and skills, i.e. of human capital. Three levels are considered in this chapter: countries and regions, enterprises and individual. At macro level, approaches and models used to measure the contribution of education, training and skills to economic growth are presented and discussed. The perspective is extended by approaches to evaluating the wider social benefits of education and training such as trust, citizenship, social cohesion and social capital. Then, a brief outline of approaches to investigating the significance of education and training for company performance is presented. Chapter 2 is completed by a review of methods to investigate the benefits – material and non-material – of education, training and skills for individuals. These include a discussion of the returns on education and training as well as research of benefits in an individual life course and biographical perspective.

Table of contents

1. Programme evaluation approaches and methods	51
1.1. Delimiting the evaluation scope	51
1.1.1. 'Objects' of VET evaluation	51
1.1.2. Effects and outcomes to investigate	52
1.2. Qualitative evaluation methods	54
1.2.1. Methods of qualitative inquiry	54
1.2.1.1. Data collection	54
1.2.1.2. Design of qualitative inquiry	56
1.2.1.3. Analysis strategies	56
1.2.2. Qualitative inquiry in evaluation	57
1.2.2.1. Evaluating individual outcomes	57
1.2.2.2. Process studies: looking at how things happen	57
1.2.2.3. Implementation evaluation	58
1.2.2.4. Studying logic models and theories of action	58
1.2.2.5. Evaluability assessment	59
1.3. Quantitative evaluation methods	59
1.3.1. Fundamental evaluation problems	59
1.3.1.1. Distinction between programme and other effects	60
1.3.1.2. Determining the counterfactual	60
1.3.1.3. Short- and long-term outcomes	60
1.3.2. Selectivity, heterogeneity and validity	61
1.3.2.1. Selection bias	61
1.3.2.2. Heterogeneity	61
1.3.2.3. Internal and external validity	62
1.3.3. Evaluation methods at micro level	62
1.3.3.1. Social experiments	63
1.3.3.2. Quasi-experimental designs: the matching approach	65
1.3.3.3. 'Before-after' comparison	66
1.3.3.4. The difference-in-differences method	67
1.3.3.5. Cross section method	67
1.3.4. Evaluation at aggregate level	68
1.3.4.1. Indirect effects	68
1.3.4.2. Cost-benefit analysis	69
1.3.5. Choice of methods and data requirements	71
1.3.5.1. Choice of method	71
1.3.5.2. Data requirements	71
1.4. Discussion	72
1.4.1. Logic and complementarity of qualitative and quantitative methods	72
1.4.2. Evaluation process and basic methodologies	73

2. Impact research: approaches and methods	74
2.1. Social and economic benefits of education and training: basic approaches	74
2.1.1. Modelling the determinants of economic growth and productivity	76
2.1.1.1. The neo-classical model: growth accounting	77
2.1.1.2. New theories on endogenous growth	77
2.1.1.3. Methodological reservations	79
2.1.2. Addressing the wider social benefits and social capital	80
2.1.2.1. Externalities, spillovers and social rates of return	80
2.1.2.2. Indirect effects of human capital on growth	81
2.1.2.3. Social infrastructure and social capital	81
2.1.2.4. Techniques to evaluate ‘macrosocial’ benefits of education and training	82
2.2. Methods to investigate the impact of skills on company performance	85
2.2.1. Training investment and company performance	86
2.2.2. Enterprise surveys	86
2.2.3. Weaknesses in defining training	87
2.3. Investigating individual benefits of education and training	88
2.3.1. Private returns on investment in human capital	88
2.3.1.1. Rates of return	88
2.3.1.2. Methodological challenges	90
2.3.2. Benefits of education and training in the life course	90
2.3.2.1. Life-course research	91
2.3.2.2. Sources of longitudinal data	96
2.3.2.3. Individual biography	98
2.4. Discussion	100
2.4.1. Education, training and economic growth	100
2.4.2. Social benefits and social infrastructure	100
2.4.3. Impact on company performance	101
2.4.4. Individual benefits and life-course research	101

List of tables, figures and boxes

Tables

2.1.	Typical research designs for impact evaluations	63
2.2.	Comparison of different evaluation methods	64

Figures

2.1.	Ashenfelter's dip	66
2.2.	Benefits of education and training	75
2.3.	The course of two cohorts through education/training and employment	95
2.4.	Unemployment rates in OECD countries 1965-2003	96

Boxes

2.1.	Terms used to characterise the results of interventions	53
2.2.	Internet, e-mail and discussion groups on qualitative methods	58
2.3.	Programme effects at aggregate level	69
2.4.	GDP and GNP as indicators for economic growth	76
2.5.	Defining human capital	76
2.6.	The neo-classical model: growth accounting	77
2.7.	Social infrastructure	82
2.8.	Defining social capital	83
2.9.	Ecological fallacy	84
2.10.	Productivity, profitability and market value – some definitions	85
2.11.	Some definitions of returns and costs	88
2.12.	Calculation of rates of return	89
2.13.	The cohort concept	92
2.14.	Age, cohort and period effects – a fictitious example	97
2.15.	National data sets for longitudinal analyses of the impacts of education and training	98
2.16.	The European Community Household Panel (ECHP)	99

1. Programme evaluation approaches and methods

This section deals with methodologies used in programme evaluation. ‘Programme’ includes diverse measures and interventions of a limited duration, aiming to solve specific problems, to achieve specific objectives and to be targeted to specific groups. Programmes can be implemented at local, regional, national or international/EU level and are not designed to interact with other interventions or institutions (e.g. employment or education and training). Programmes are distinct from the more general policies targeted at human capital development in a socioeconomic framework. Typical examples are training programmes and reintegration measures developed in the framework of active labour-market policies which aim to combat unemployment and other forms of exclusion.

Before deciding on the most appropriate method(s) to be applied to an evaluation, one must clearly define the objects of evaluation and the main characteristics of the intervention under investigation. In agreement with commissioners and stakeholders, it is necessary to define the kind and level of effects and outcomes that will be investigated according to the goals and objectives of the programme or policies to be evaluated. Following this, the evaluator has to consider the most appropriate evaluation design(s), method(s) and data requirements.

Programme evaluation can be approached both through qualitative and quantitative methods. Rather than alternatives, they are complementary. In this chapter, we will first discuss in detail the designs, strategies and tools developed in both methods before discussing how they can be applied in a complementary manner to the evaluation of VET programmes.

1.1. Delimiting the evaluation scope

1.1.1. ‘Objects’ of VET evaluation

The objects of evaluation (or ‘evaluands’), can be described according to several dimensions: domain, character and target of the intervention. These mostly overlap and can be broken down further.

‘Domain’ refers to the field or contextual demarcation of an intervention. Evaluation in the domain of VET may, for example, specifically refer to initial VET and to its role in facilitating the transition from school to work. Other domains of VET evaluation are continuing vocational training, vocational guidance, placement, workplace learning, lifelong learning, non-formal learning, active labour-market policies, etc.

The ‘character’ of an intervention denotes the type of a measure. A measure can be a (pilot) project, a programme, a policy or a reform. Examples are pilot projects implementing a new curriculum in a limited number of schools, programmes to impart basic skills, active labour-market policies, reform of initial vocational education, etc. (!).

The ‘target’ can be individuals or groups with particular characteristics (e.g. young people, immigrants), a (public or private) organisation or part of it, a region, an economic sector, a country or a cluster of countries (e.g. EU, Nordic countries).

Evaluations will normally cover and combine all three dimensions, taking into account the objectives of the intervention. It is thus necessary to properly define the domain, character and target of a measure and respectively its evaluation. As an example: ‘the objective of this pilot project is to enhance employability of young immigrants. It focuses on initial basic training and language learning including vocational guidance and will be carried out by in-company training in the manufacturing sector in the region XY, complemented by school courses.’

(!) Part 3 of this report discusses evaluations of education and training systems, reforms and programmes.

When defining the objects of evaluation, it is necessary to consider the following characteristics ⁽²⁾:

- (a) boundedness of settings: isolated evaluation objects are increasingly rare. On the contrary, they are often configurations, embedded in specific contexts. Typically, a VET measure is bundled together with a package of incentives, vocational guidance measures and qualifications and therefore has to be evaluated in this wider context;
- (b) input characteristics: programmes are operationalised through policy instruments. In VET, common instruments would include new forms of funding, new curricula or new training courses. Inputs in many VET programmes are also carefully tailored to individual, local or enterprise needs. This has consequences for sampling and the scale of evaluation. Input characteristics will have implications for the potential to generalise on the basis of evaluation findings;
- (c) the immediate context: the context within which an input is located can be relatively standardised or diverse. Context may vary according to the characteristic of a geographical area or in terms of the institutional setting within which programmes are delivered or policies are expected to have an impact. In VET, therefore, the relevant immediate context may be the labour market, a training provider or an enterprise. A highly diversified initiative may be implemented across different kinds of contexts and even within a single context there may be considerable variety;
- (d) modes of delivery: the same input or instrument can be delivered in different ways. For example, a needs analysis may be undertaken through a local survey as part of the recruitment process of potential trainees or by a company reanalysing its personnel data. Nowadays it would be

more common for programmes to be delivered through partnership arrangements rather than through a single administrative chain. This will often be the case, for example, in VET measures delivered through EU Structural Funds;

- (e) number of stakeholders: stakeholders are those who have an interest in the evaluation and what is being evaluated. Within decentralised, multi-agency programmes there are often many stakeholders, each with their own evaluation questions and judgement criteria. These might include regional authorities, training providers, commissioners, sector representatives, social partners and European institutions;
- (f) the degree of consensus: policies and programmes may generate various degrees of consensus among stakeholders. When there is a high degree of consensus about an evaluation, it will be possible to apply agreed criteria easily. Where there is lower consensus, quite different criteria may need to be applied and more work may be needed to bring together different interests and perspectives. This shapes not only methodology but also the work required of evaluators.

1.1.2. Effects and outcomes to investigate ⁽³⁾

Evaluation investigates the results of an intervention. Characterising these results, the distinction between effects, outcomes, output, impact, etc., is not always clear. Even more, they refer to different notions in different languages. Box 2.1 provides some definitions.

The outcomes of a programme or policy may constitute benefits at different levels:

- (a) outcomes for individual programme participants may be economic or work-related, such as: earnings, job search efficiency, employment and employment stability, job satisfaction and productivity, occupational fits and status. Or, they may refer to advancement in learning, skills and competences (including 'higher-order

⁽²⁾ The following section draws mainly on Stern (2004).

⁽³⁾ This section draws mainly on Grubb and Ryan (1999: 32 ff.).

Box 2.1. Terms used to characterise the results of interventions

Effect:	socioeconomic change resulting directly or indirectly from an intervention. Effects include the results and impacts of an intervention, whether positive or negative, intended or not.
Output:	immediate and direct tangible result of an intervention.
Outcome:	the positive or negative longer term socioeconomic changes or impacts that occur directly or indirectly as a result of an intervention's inputs, activities and outputs.
Impact:	a general term used to describe the effects of a programme, policy or socioeconomic change. Impacts can be positive or negative, direct or indirect as well as foreseen or unforeseen.

NB: See the glossary for the various sources and the translation of the terms in French and German.
Source: Extracted from the glossary in Annex 1.

skills/competences' like problem solving or other key skills/competences) and to non-material benefits such as better health, parenting, self-fulfilment, occupational identity, crime reduction, etc.

- (b) outcomes for other agents include benefits for other individuals or employers/companies (e.g. increased productivity) although these are difficult to measure ⁽⁴⁾;
- (c) outcomes for welfare and social measures include, for example, reduced time in welfare measures or unemployment resulting from more efficient job searching or upgrading of skills;
- (d) social and economic outcomes refer to, for example, an increase in economic growth, competitiveness and innovativeness of a country. They could also involve a decrease in criminality or drug use (e.g. via increased unemployment), tolerance, reduction in teenage pregnancy and improved health status of population, citizenship and social cohesion; in

economics these outcomes are known as 'externalities' or 'spillover effects'.

Outcomes of a programme might be positive or negative, direct or indirect, intended or not. For example, participation in a training programme may increase directly the employment opportunities of the participants, but reduce those of non-participants.

In analysing the effects of a programme one has also to consider their duration and heterogeneity. First, many programmes yield benefits that tend to decrease gradually in the medium term and become insignificant after some years. For example, programmes of job-search assistance may prove cost-effective in the short run, but poor in the longer run compared to intensive education and training programmes which yield better long-term results ⁽⁵⁾. Second, programme effects usually refer to average effects; however, there may be considerable variations across participants (marginal effects). Programmes are normally less effective for the best-prepared and most able participants (who can easily find a job on their own) as well as for the most disadvantaged (who have serious problems which lie beyond the scope of a programme and thus can hardly be resolved by it). Interventions are likely to be most effective for the group in the middle. Another aspect is the variation of effects with size of the programme. For example, in the US in most cases only a fraction of those who are eligible do participate, whereas in Europe programmes typically cater for the majority of potential beneficiaries, as with active labour-market programmes ⁽⁶⁾.

One of the fundamental evaluation problems is the question – which is in principle irresolvable – what would otherwise have been, i.e. what would be the situation if a programme or reform had not been implemented? This 'counter-factual' aspect refers to all levels (individuals, employers, public) and categories, and to all stages of the

⁽⁴⁾ Cf. Part 4, Chapter 3 and the contribution of Hansson et al. to the background report (2004).

⁽⁵⁾ The competences for finding a job as imparted in job-search programmes are not necessarily the same as those for keeping a job. The skills required to keep a job are rarely incorporated in programmes and thus in evaluations, because they require a long term perspective.

⁽⁶⁾ This issue of economics of scale in terms of programme size is still under-researched.

intervention. It is not possible to judge the particular effects of an intervention without comparing it with a situation or people without intervention; nor is it possible to observe simultaneously a situation or outcome both with and without having carried out a reform, programme or project.

Other major evaluation difficulties when investigating outcomes and effects are:

- (a) the distinction between programme/reform effects and effects induced by other factors of influence;
- (a) the selectivity of participants in a programme or target groups of a reform;
- (a) the internal and external validity of evaluations ⁽⁷⁾.

1.2. Qualitative evaluation methods ⁽⁸⁾

Qualitative methods ‘tell the programme’s story by capturing and communicating the participants stories’ (Patton, 2002: 10). Qualitative methods are often used in evaluations, alone or in combination with quantitative methods, because they are able to capture specific types of information that other methods fail to bring to the surface. They deepen understanding of the process and outcomes of a programme. They are used in evaluation to improve programme design and inform decision-making. Therefore, the primary audience for qualitative evaluation results are the programme financiers and administrators, staff and participants.

Though they can be used on their own, qualitative evaluation methods often usefully complement quantitative methods. While the latter are used to generalise results and to make them comparable by using standardised variables and common dimensions, the former intend to increase the depth of understanding beyond quantitative data. The two methods are not mutually exclusive but complementary strategies for research and evaluation.

In an attempt to characterise in a few

sentences the nature of qualitative research in comparison to quantitative methods (with the necessary generalisation that this implies), Denzin and Lincoln (2000: 8) write: ‘The word “qualitative” implies an emphasis on the qualities of entities and on processes and meanings that are not experimentally examined or measured (if measured at all) in terms of quantity, amount, intensity, or frequency.’

Qualitative inquiry intends to offer the full picture of the ways an intervention does perform (or not) and allows for capturing non measurable effects. It can be used in a formative and summative perspective, allowing in both cases deeper understanding of processes, dynamics and individual cases at stake. It is a tool for programme improvement.

Smith (1986: 23, cited in Guba and Lincoln, 1989: 159) summarises the qualitative approach as the ‘long-term and first-hand study of a case by the investigator for the purpose of understanding and describing human actions in the context of that case. Fields methods are used to collect data, including direct observation of action in its natural context, clinical interviews to elicit the multiple leanings of participants in that case, and the collection of documents. A qualitative approach leads to reports primarily in the form of words, pictures and display rather than formal models or statistical findings.’

In this section, we will review and outline some of the major designs and methods of what Patton calls ‘qualitative inquiry’, as well as the way it is used in evaluation.

1.2.1. Methods of qualitative inquiry

The added-value of qualitative inquiry is the wealth of detailed information provided on a limited number of people and cases.

1.2.1.1. Data collection

Fieldwork is the central activity of qualitative inquiry. Collecting qualitative data requires direct personal experience and engagement ⁽⁹⁾ which makes it possible to describe and

⁽⁷⁾ See for more details Section 1.3.2.3 and the subsequent sections for methods to solve evaluation problems at micro level.

⁽⁸⁾ This section is mainly based on Patton (2002).

⁽⁹⁾ Sometimes the evaluator even interacts with the situation under study, as a participant observer.

understand both externally observable behaviour and internal states (worldviews, opinions, values, attitudes and symbolic constructs; Patton, 2002: 48). This kind of approach allows for perceiving impacts that are not measurable or not measured by standardised tests.

In qualitative inquiry, the human being is the instrument of data collection. Therefore, the researcher's personal experience and insights are an important part of the inquiry and critical understanding of the phenomenon. However, neutrality remains an ideal that the researcher must strive to attain to avoid bias and false interpretations. Achieving neutrality requires that the investigator carefully reflects on, deals with and reports on potential sources of bias and errors.

'Qualitative evaluation researchers [...] conceive a programme as dynamic and developing, with "treatments" changing in subtle but important ways as staff learns what does and does not work, as clients move in and out, and as conditions of delivery are altered.' The primary challenge for evaluators is 'describing and understanding these dynamic programme processes and their holistic effects on participants so as to provide information for programme improvement' (Patton, 2002: 54). Consequently the data collection pays particular attention to processes. The evaluator assumes that change is ongoing and focuses on system and situation dynamics.

While quantitative measures are succinct, standardised and easily aggregated for analysis and comparison, qualitative findings are more lengthy, detailed and variable in content. Their analysis is difficult because responses are neither systematic nor standardised.

Four main kinds of data collection tool are used in qualitative inquiry (Patton, 2002: 4, 20-28).

First are questionnaires with open-ended questions; these permit understanding and capturing the points of view of respondents without predetermining these points of view through prior selection of questionnaire categories. Data consist of direct quotations.

Responses can be analysed and patterns defined by text analyses. Responses are the most elementary form of qualitative data. They find their limitation in the writing skills of respondents and they often require a considerable investment in terms of time and efforts by the person completing the questionnaire. In addition, probing of responses is impossible.

Second are interviews. Using open-ended questions and probes, the evaluator collects in depth responses about people's experiences, perceptions, opinions, feelings and knowledge. Programme evaluation interviews aim to capture the perspective of programme participants, staff and others associated with the programme. Data consist of verbatim quotations with sufficient context to be interpretable. The limitations of interviews are linked to the potential influence the interviewer actually exerts on the answers of the interviewee, through the choice of questions, body language, interruptions, etc. (Kvale, 2003: 231).

Next come observations. The purpose of observational analysis is to take the reader into the setting that was observed. Observations provide fieldwork descriptions of activities, behaviours, actions, conversations, interpersonal interactions, organisational or community processes, etc. Data consist of field notes, which should be factual and accurate without being too detailed. Field notes should also include information on the context of the observation. However, observations are highly labour intensive and, therefore, constitute a relatively expensive means for data collection. Also, the evaluator must always take into account the influence his or her presence may have on the situation.

Finally there are documents. The evaluator consults written documents from organisational or programme records, official publications and reports, written responses to open-ended surveys, etc. Data consist of excerpts of the documents captured in a way that records and preserves context.

In conclusion, qualitative inquiry data typically come from fieldwork. 'They take the reader into the time and place of observation

[and] capture and communicate someone else's experience of the world in his or her own words' (Patton, 2002: 47).

After data collection, the task of the evaluator is to transform these raw data into readable narrative descriptions with major themes, categories and illustrative case examples extracted through content analysis. Although there are methods that should be followed while collecting qualitative data, its validity, reliability and representativity will depend a great deal on the methodological skills, the sensitivity and the integrity of the researcher.

1.2.1.2. *Design of qualitative inquiry*

Three main designs are characteristic of qualitative inquiry (Patton, 2002: 39-47):

- (a) naturalistic inquiry: the evaluator studies real-world situations as they unfold naturally. He/she does not exert any manipulation or control over the situation under analysis and he/she is open to whatever emerged (i.e. no predetermined constraints on findings);
- (b) emergent flexibility: characterises the evaluator's readiness to pursue new paths as they emerge and his/her openness to adapt the inquiry as understanding deepens and/or situation changes;
- (c) purposeful sampling: the very nature of qualitative inquiry requires a sampling strategy that will lead to the study of a relatively small number of special cases that are particularly rich information cases for the programme under evaluation. What matters is the large amount and the depth of data that is collected on these specific cases, rather than the number of cases. At the extreme, the sample can be constituted by a single case.

1.2.1.3. *Analysis strategies* ⁽¹⁰⁾

A number of analysis strategies can be identified.

- (a) Case orientation is a feature of qualitative studies. A case can be a person, an event, a programme, an organisation, a time period, a critical incident or a community. When analysing case studies, the evaluator must be true to, respect, and capture the details of the individual case. Cross-case analysis depends on the quality and comparability of individual case studies. Case studies are particularly valuable in programme evaluation when the evaluator needs to understand variations among individual participants or programmes.
- (b) Inductive analysis consists of adopting a discovery rather than a verification approach (Guba and Lincoln, 1989: 59). This means that enquiry does not start on the basis of a preformed hypothesis to be verified. On the contrary, there are no preconceptions towards the nature of findings. Enquiry begins with specific observations and builds up towards extracting general patterns. Categories or dimensions of analysis emerge from open-ended observations as the enquirer comes to understand patterns that exist in the phenomenon being investigated. The evaluator does not presuppose in advance what the important dimensions will be ⁽¹¹⁾. An investigation may start from inductive qualitative approach to find out about important questions and variables (exploratory work). Then, it may move on to deductive hypothesis testing or outcome measurement to confirm and/or generalise exploratory conclusions drawn from the first stage. It can then become inductive again, to look for rival hypothesis and unanticipated or unmeasured factors that can be then in

⁽¹⁰⁾ See Patton (2002: 55-66).

⁽¹¹⁾ On the contrary, hypothetical deductive experimental research requires specification of the main variables and hypothesis before data collection begins. A clear example of the difference between the two approaches is the contrast between multiple choice questionnaire (MCQ) and open-ended questions. The MCQ is deductive; the categories are based on theory or predefined criteria (e.g. programme goals). The open-ended questionnaire permits respondents to describe what is meaningful and salient without being driven towards standardised categories.

turn tested on a broader scale and generalised ⁽¹²⁾.

- (c) Adopting a holistic perspective is striving to understand a programme as a whole, i.e. a complex system that is more than the sum of its parts. ‘3 sticks are just 3 sticks until one places them to form a triangle’. The analysis therefore focuses on complex interdependencies and system dynamics ⁽¹³⁾.
- (d) Context sensitivity means that in qualitative enquiry the context is considered as critical to understanding. It is crucial to place findings in their social, historical, and sequential context. At the same time, one should be careful about ‘decontextualised’ generalisations ⁽¹⁴⁾ across time and space but instead emphasise comparative case analysis and extrapolate patterns for possible transferability and adaptation in new settings (Patton, 2002: 41) ⁽¹⁵⁾. According to Guba and Lincoln (1989: 36-37), ‘context-stripping’, i.e. evaluating an intervention trying to control specific contextual conditions, is one of the reasons why evaluations are often found irrelevant at local level. They suggest (p. 60) taking into account the contextual factors rather than trying to control them physically or statistically. According to them, an excessive preoccupation with internal validity reduces the likelihood of results being externally valid ⁽¹⁶⁾. They conclude (p. 70) that ‘it may be possible to assess the evaluand’s merit in a general way, but one cannot assess its worth except in the individual case’.
- (e) Reflexivity is important as well. The evaluator needs to be particularly aware of the culturally-bounded frameworks he/she is bringing with him, i.e. the filters that help

him to interpret the world ⁽¹⁷⁾. Complete objectivity being impossible and pure subjectivity undermining credibility, the evaluator’s focus becomes understanding and depicting the world authentically in all its complexity while being self-analytical, politically aware, and reflexive (Patton, 2002: 41).

1.2.2. Qualitative inquiry in evaluation ⁽¹⁸⁾

‘The failure to find statistically significant difference in comparing people on some outcome measures does not mean that there are no important differences among these people on those outcomes. The difference may simply be qualitative rather than quantitative’ (Patton, 2002: 151) ⁽¹⁹⁾.

In this section, we will review some typical evaluation cases where qualitative inquiry is particularly suitable.

1.2.2.1. Evaluating individual outcomes

Qualitative studies offer a method for capturing and reporting on individual outcomes. This is particularly interesting when interventions foresee individualised approaches because they rely on the assumption that outcomes will be qualitatively different for different clients ⁽²⁰⁾. A similar situation applies when common activities may have drastically different outcomes given the initial profile and background of participants.

1.2.2.2. Process studies: looking at how things happen

Process evaluation aims at elucidating and understanding the internal dynamics of how programmes, organisations or relationships operate. Qualitative inquiry is appropriate because:

⁽¹²⁾ The ‘grounded-theory’ approach developed by Glaser and Strauss (1999) is based on this mode of analysis.

⁽¹³⁾ See Part 3, Section 2.3 on systemic analysis applied to VET systems.

⁽¹⁴⁾ On the contrary, the purpose of quantitative approaches is to decontextualise the findings.

⁽¹⁵⁾ For the application of this strategy to VET system evaluation see the discussions in Part 3, Sections 2.4.3 and 3.3.

⁽¹⁶⁾ By controlling contextual factors of influence, researchers can explain better the relationship between the variables under of observation, thus increasing internal validity. However, an excessive control of the context reduces the solidity of the findings in real world situation, thus reducing external validity. See Section 1.3.2.3 on internal and external validity.

⁽¹⁷⁾ See Part 3, Section 2.5.2.1 for common culturally-bounded frameworks in VET.

⁽¹⁸⁾ This section is based on Patton (2002: 151–168) on evaluation applications for qualitative inquiry.

⁽¹⁹⁾ In quantitative approaches, these differences are unobserved or considered unobservable.

⁽²⁰⁾ Flexibility, adaptability and individualisation can be important to the effectiveness of educational and human services. In this case, quantitative measures of outcomes are particularly inappropriate.

- (a) depicting process requires detailed description of how people engage with each other;
- (b) experience of process typically varies for different people so it needs to be captured 'in their own words';
- (c) process is fluid and dynamic and cannot therefore be summarised on a rating scale at one point in time;
- (d) participants' perceptions are key process considerations.

1.2.2.3. *Implementation evaluation*

The goal of implementation evaluation ⁽²¹⁾ is to make sure that a policy is put into operation according to its design. Unless this is known, there is little reason to expect the desired outcomes. In addition, without knowledge of implementation, the decision-maker lacks information about what produced the (lack of) desired outcomes.

In addition, implementation at local level seldom follows exactly the proposed design. There is a degree of adaptation to local needs and situations that is desirable. Qualitative descriptions are needed to capture these diversities and contrasts as well as how they are related to the outcomes.

Implementation evaluation tells decision-makers what is going on in the programme, how the programme has developed and how and why it deviated from initial plans and expectations. Qualitative methods are ideally suited to this kind of evaluation ⁽²²⁾.

1.2.2.4. *Studying logic models and theories of action*

Logic models and theories of action ⁽²³⁾ describe the connections between programme input, activities and process (implementation), outputs, immediate outcomes and long-term impact ⁽²⁴⁾. Patton uses the following example of the DARE (drug abuse resistance education) programme (2002: 162-163). It followed a simple logic model: recruit and train police

Box 2.2. **Internet, e-mail and discussion groups on qualitative methods**

QUALRS-L@listserv.uga.edu

qualitative research for the human science

QUALNET@listserv.bc.edu

qualitative research in management and organisation studies

QUAL-L@scu.edu.au

qualitative research list, initiated by Penn State, but immediately attracted a broader audience

Other resources for qualitative evaluation and research:

EVALTALK@bama.ua.edu

American Evaluation Association (AEA) discussion list

www.eval.org

AEA home page with links to evaluation organisations, training programmes and Internet resources

METHODS@cios.org

a list for social science research methods instructors

NB: These sites may change, and this list is not exhaustive. It is meant to be suggestive of the qualitative resources available through the Internet.

Source: Patton, 2002: 29.

officers (input) to teach children the dangers of drug use; have the police officers teach, in uniform, children in special classes in school (implementation); as a result, the children will find the teacher credible (process evaluation) and learn facts about drugs (cognitive outcome); which will convince them not to use drugs (attitude change outcome); which will result in students not using drugs (behaviour change outcome), which will ultimately show up in community indicators showing less drug use (impact). However, this theory of action did not work in practice ⁽²⁵⁾.

To find out the reasons for failure, evaluation focuses on the difference between the logic model and theory of action and what happens in reality when it is implemented. Interviewing participants and front-line staff reveals the theory-in-use and helps explain the reasons for, and implications of, discrepancies.

⁽²¹⁾ See also Section 1.1 for stages and phases of evaluation.

⁽²²⁾ See also discussion on 'opening the black-box' in Part 3 Section 4.4.2.

⁽²³⁾ See also Part 1, Section 2.1.3: Action-theory.

⁽²⁴⁾ Also called 'intervention theory' or 'programme theory' (Vedung, 2000b: 138-139).

⁽²⁵⁾ See Part 1, Table 1.1 for a description of the four stages of the logic model of VET programmes.

1.2.2.5. *Evaluability assessment*

Evaluability assessments determine whether a programme is sufficiently conceptualised and consistently implemented to undertake a formal, rigorous summative evaluation aimed at determining overall effectiveness. They aim to ensure that the intervention or model are clearly identifiable and logical, that outcomes are clear, specific and evaluable, and that implementation strategies are reasonably and logically related to expected outcomes. They are conducted through interviews, document analysis and observations.

1.3. Quantitative evaluation methods

Quantitative evaluation methods are mainly used in summative evaluations of programmes or interventions. The objective is to measure outcomes and impact of a specified (independent) variable, such as training, and to compare these with a situation without such a programme or with people not participating in that programme. Quantitative methods thus serve to provide appropriate information for decision-makers on accountability, efficiency or effectiveness of an intervention, ideally by controlling for (i.e. eliminating) all other factors of influence, such as social background, gender or ethnicity, which are not or only loosely related to the independent variable.

Having defined, in the first stage of an intervention, the objectives and design of the programme, its evaluation requires the definition of criteria on how to judge its success. Evaluating programme impact requires ‘disentangling the effects of the programme from other exogenous factors.’ (Hujer et al., 2004a: Section 2.1.4).

Success criteria could be, at micro level: earnings after participation; unemployment risks; acquisition of skills and competences; employment opportunities and mobility; self-consciousness; productivity at work; political and social participation/inclusion; etc. Each of

these criteria has to be defined clearly in light of the programme objectives. For example, reduced unemployment risks could be the result of training in more efficient job search behaviour; higher productivity at work could be achieved by management training courses, and the like. Thus, outcomes, effects or impacts refer to the – expectedly positive – change of these criteria compared with the situation before the programme or with a comparison group.

At aggregate level, effects outside the area covered by a programme or policy have to be taken into account. These refer mainly to indirect effects (spillovers) and external effects which are of benefit to other individuals, firms or society in general.

1.3.1. **Fundamental evaluation problems**

Several fundamental evaluation problems have to be resolved at micro and aggregate level. First, methods have to be applied which ensure the proper distinction between programme effects/outcomes and effects caused by other factors. Second, the counterfactual situation has to be clarified: what would have been without the intervention? Third, different time horizons of impacts have to be considered. Finally, there are a number of potential distortions which might impair the validity of an evaluation exercise.

A major problem underlying each evaluation and impact research (see Chapter 2) is the difficulty of proving causality, i.e. the cause and effect relationships between education and training, on the one hand, and outcomes such as earnings, economic performance on the other. Normal statistical methods such as regressions/correlations or econometrics only state a relationship or association between dependent and independent variables, but not necessarily a causation. However, because an educational programme usually precedes the outcome, a causal link can be assumed. Thus, the inference of causation comes from the logic model of intervention (what causes what, and why, MacMahon, 1998: 459) ⁽²⁶⁾ and is not deduced from correlations – micro or

⁽²⁶⁾ See Part 1, Table 1.1 for the logic model of VET programmes.

macro – or from the data (except when panel longitudinal data are available).

1.3.1.1. *Distinction between programme and other effects*

One of the major problems facing evaluation is to distinguish between the outcomes caused by the programme itself and those generated by other influencing factors. Even in well-conducted evaluations using control groups or other approaches part of the outcome (e.g. earnings, employment situation) cannot be explained by the intervention itself but is the result of interaction effects of extraneous factors. If such factors are non-observable or are not observed in the data set, they will confound the mere interventionary effects. Examples are unobserved characteristics or life situations of participants such as motivation, illness or family problems, external changes in the overall socioeconomic environment, etc. Although sophisticated econometric methods have been developed, there are still residuals in explaining the causes of outcomes. Many of these causes can only be identified by qualitative methods (see Section 2.2). However, concerning unobserved characteristics, some advances have been made by studying genetically identical twins (monozygotic-one egg) and thus by eliminating, for example, the effects of innate ability and family background (Ashenfelter and Rouse, forthcoming; Ashenfelter and Krueger, 1994). However, studies in psychology on identical twins have demonstrated that the effect of the family background/environment can only partially be eliminated.

1.3.1.2. *Determining the counterfactual*

The second fundamental evaluation problem is closely connected with the question of how to distinguish between programme and other effects as mentioned above. Unfortunately the outcomes for a single individual can never be observed twice simultaneously, both with

participation of an individual in a particular programme and without participation of the same individual. Each hypothetical alternative is, of course, unobservable.

To resolve this problem, a number of approaches have been developed, particularly at micro level.

The most common is to compare the outcomes for participants (treatment group) with a control group. This, however, requires both groups to have the same structural characteristics, both observable and unobservable⁽²⁷⁾. The outcome for the control group, for example increased earnings without having participated in an intervention, is taken as the counterfactual outcome of the treatment group. Other approaches include observing participants before and after the programme or comparing the change of outcomes, such as earnings, of participants before and after the programme with the change of earnings of non-participants in the same period. These and other approaches will be discussed in more detail in Section 1.3.3.

1.3.1.3. *Short- and long-term outcomes*

Any evaluation of a programme encounters the problem that outcomes – or benefits in general – will appear in different periods: in the short-, medium- and long-term. Furthermore, costs may typically not occur in the same periods as outcomes. These problems become relevant if we reconsider the functions of an evaluation. It has (see Part 1, Section 1.3) primarily an information, allocation and legitimisation function. Information on a programme based on evaluation aims to improve the programme or to reshape similar follow-up interventions. Allocation of funds and accountability compares the money spent with the outcomes of the intervention. Finally, evaluation results may or may not legitimise policy.

These functions are mainly of short- and medium-term nature. However, benefits may also accrue in the longer term, or they may level out gradually or be reversed in the course

⁽²⁷⁾ Examples for observables are age, gender, level of qualification; unobservable characteristics might be ability, motivation, values, attitudes.

of time. In consequence, programme efficiency should, if possible, be evaluated in the short-, medium- and long-run. However, decision-making, demonstration of accountability and legitimisation are issues of the short-, or sometimes medium-term agendas of governments and stakeholders. Furthermore, it will become increasingly difficult to attribute outcomes such as changes in earnings or employment situation to a programme conducted many years ago. A large number of other intervening influences external to the intervention, will increasingly superpose the programme effects.

1.3.2. Selectivity, heterogeneity and validity ⁽²⁸⁾

1.3.2.1. Selection bias

Selection bias occurs when unobservable characteristics of participants, such as ability or motivation, influence their participation in a programme and their success after the intervention ⁽²⁹⁾. A selection bias occurs if, for example, more motivated people are more likely to participate in the programme and afterwards to be employed. If the objective of the programme is to enhance employability, the selection of participants disadvantages all those who are the main target group of the intervention but do not participate. Empirical evidence confirms a similar situation if willingness to undergo further education and training is closely associated with educational background. If a programme for continuing training or retraining of unemployed people does not take into account the educational background of participants, the outcomes might be biased. Controlling for selectivity is, therefore, a major task of any evaluation.

1.3.2.2. Heterogeneity

Another important evaluation problem may be caused by heterogeneity. Heterogeneity might arise at three levels: individual, programme and regional.

Heterogeneity of participants may occur if, for example, a programme to enhance employment prospects by further training involves the long-term as well as the short-term unemployed. The programme effects for these heterogeneous groups may differ significantly, they might be positive for the short-term unemployed and negative for the long-term unemployed. Taking the average of both groups will lead to a biased estimation of the treatment effect. The effect for the short-term unemployed is underestimated and for the long-term unemployed overestimated. Clearly, this bias can be avoided by estimating separately the effects for both groups.

Programme heterogeneity happens when programmes with similar objectives but with different duration, reach, content and effects are compared. For example, comparing a two-week computer course with a six-month course on informatics will be misleading. Estimating the effect by taking the average over both programmes thus will give a false indication of the 'true' effects.

Finally, heterogeneity in terms of the geographical location of a programme might bias its outcomes. An example would be an employment programme run both in a booming region and a depressed region. The programme might work well in the booming region but not in the other region. Taking together the programme effects will cause a bias similar to those mentioned above ⁽³⁰⁾. This bias can be avoided if data allow for regional analyses.

Disregarding the heterogeneity of participants, programmes and regions leads to an underestimation (or overestimation) of the true effect. Clearly this is an unnecessary source of bias which has to be avoided. Therefore, evaluations should be carried out separately with respect to major individuals' characteristics, regional aspects and programme differences.

⁽²⁸⁾ This section draws on Hujer et al. (2004a, Sections 4.1 and 4.2).

⁽²⁹⁾ Selection bias might also occur when choosing a control group (Section 1.3.1.2).

⁽³⁰⁾ Similar problems arise when participants are located in a booming region, and the control group members come mainly from a depressed region.

1.3.2.3. *Internal and external validity*

The concept of validity refers to the quality of evaluation (Cohen and Manion, 1994: 170 ff.).

Internal validity refers to the accuracy of the measures and results obtained. It is achieved when the design and method of evaluation avoid biases such as selectivity and heterogeneity (see above) and if high quality data as well as suited methods are used. Lack of internal validity results from the effect from unobservable variables other than the treatment, from unreliable instruments and methods leading to errors. Conclusions on programme effects will be more valid if all specific aspects as well as contexts of an intervention are considered. Therefore, it is necessary to control strictly the relevant parameters which influence programme effects.

External validity is achieved if the conclusions on programme effects can be generalised and transferred to other contexts such as other groups, regions, subjects or time periods. Strong external validity allows transfer of lessons learned from an intervention to similar activities, or generalisation of features of best practice. External validity may be hampered by a lack of representativity of target populations, inadequate operationalisation of the independent and dependent variables, and by interaction effects of extraneous factors.

1.3.3. **Evaluation methods at micro level** ⁽³¹⁾

A proper evaluation design for the quantitative measurement of programme outcomes and effects has to consider at least four methodological issues (Stockmann, 1997: 107, based on Rossi and Freeman, 1988: 212 f.):

- (a) type of study and database, for example experiments, retrospective studies; comparison with control group; expert judgements, etc.;
- (b) choice of target units, for example random or non-random assignments; uncontrolled selection;

- (c) type of control, for example constructed or statistical control; reflexive;
- (d) collection points of outcome data, for example before or after an intervention; or retrospectively.

Research designs intend to investigate planned and unplanned outcomes of a programme and to identify their causal relations. However, they should not only be restricted to direct outcomes of the programme itself but include also outcomes that occur from other causes found in the social context where a programme is embedded. Whereas most programmes allow for the measurement of direct outcomes, problems of self-selection, high complexity of the phenomena surrounding a programme and confounding effects of the intervention with effects produced by other variables make it difficult to assess all outcomes, including intangible ones.

Table 2.1 summarises the most usual evaluation designs and approaches (first column) ⁽³²⁾. The second column indicates the way the target (treatment) group is selected and whether or not this selection is controlled by the researcher. The third column points to the way outcomes are assessed – by comparison with a control group – by comparing the situation of the target group before and after a programme, or by comparing outcomes with norms or expectations. The last column indicates the points in time where data on outcomes are collected.

Table 2.2 summarises the different evaluation methods presented below with short descriptions and indicating their pros and cons. All methods try to measure, in one way or another, the outcome of programmes for programme participants. This measurement is always done by comparison with an alternative situation without such a programme.

A numerical and hypothetical example of the differences in evaluation results depending

⁽³¹⁾ This section draws on Hujer et al. (2004a: Section 2.2).

⁽³²⁾ Several of them will be discussed in more detail below. For a more comprehensive and formal description of quantitative methods at microeconomic level see Hujer et al. (2004a).

Table 2.1. **Typical research designs for impact evaluations**

Research design	Choice of target group	Type of control	Outcome data collection points ^(a)
1. Social experiments (with randomisation)	Random assignment of treatment and control group	Randomisation ^(b) often with statistical controls ^(c)	Minimum: after programme. Usually before and after; often many measures during the programme
2. Matching (quasi-experiments)	Uncontrolled selection with non-random assignment	Constructed ^(d) and/or statistical control	Minimum: after programme. Usually before and after; often many measures during the programme
3. Before-after studies	Uncontrolled selection	Reflexive control ^(e)	Minimum – before and after programme
4. Difference-in-differences studies	Uncontrolled selection	Reflexive and statistical control	Minimum – before and after programme
5. Retrospective before-after studies	Uncontrolled selection	Retrospective reflexive control	After programme with retrospective measures of before state
6. Panel studies	Uncontrolled selection	Reflexive control	More than two measures during the programme
7. Time series	Uncontrolled selection	Reflexive control	Many measures before and after the programme
8. Cross-section methods	Uncontrolled selection	Statistical control	After programme only
9. Expert-judgements model	Uncontrolled selection	Generic ^(f) and/or shadow ^(g) control	After programme only

- ^(a) The term ‘programme’ includes all kinds of interventions (see Chapter 1); ‘outcome’ or ‘effects’ include all kinds of impacts measured by the evaluation, including counterfactual outcomes (see Section 1.1.2). Examples for outcome categories are: earnings, employment situation, skills, status, and other measurable or observable criteria which relate to the programme objectives.
- ^(b) Randomised controls: target groups are divided randomly into a treatment group and a control group. The planned intervention/ programme takes place in the treatment group only.
- ^(c) Statistical controls: participating and non-participating individuals are compared by attempting to hold statistically constant differences in the composition of the two groups.
- ^(d) Constructed controls: for every individual in the treatment group, one seeks an equivalent partner (a match) which is not subject to the programme stimulus. By this parallelisation, a constructed (but nevertheless actually existing) control group for comparisons with the treatment group is created.
- ^(e) Reflexive controls: the treatment group becomes its own target population by comparing the measurements, observed before and after the intervention, of the participating individuals. In this way, differences, such as the demographic composition of the treatment group, are excluded.
- ^(f) Generic controls: intervention effects for the target group are determined by means of a comparison with established or recognised norms for typical states or changes in the target population.
- ^(g) Shadow controls: the effects on the target group subject to the programme are compared to what would have been expected by experts, programme directors, and/or participants without the programme.

Source: Stockmann, 1997: 107 (based largely on Rossi and Freeman, 1988: 212-213).

on the method used is presented in Annex 2, based on Hujer et al. (2004a).

1.3.3.1. *Social experiments*

Social experiments are based on positivist thinking ⁽³³⁾, assuming that it should be possible to measure objectively the net effects or outcomes of an intervention and to

establish causality relationships. This thinking is similar to natural sciences which intend to measure causes and effects under conditions controlled by the researcher, for example in a laboratory. Causes-impact relationships are, however, difficult to prove in almost all spheres of social life because of numerous intervening variables. Furthermore, the selection of people

⁽³³⁾ See Part 1, Section 2.1.1.

Table 2.2. Comparison of different evaluation methods

Method	Description	Advantages (+) and disadvantages (-)
Social experiment	Compares the outcome of a programme for participants and non-participants, selected randomly before the programme takes place.	<ul style="list-style-type: none"> + full control, unbiased selection and estimation of 'true' outcomes + unobservable characteristics become irrelevant - exclusion of eligible persons from programmes is problematic for ethical and legal reasons - uncontrollable interaction and moves between treatment and control group that bias results
Matching (quasi-experiments)	Compares the outcome of participants in the period after the programme, with the outcome of matched (statistical twins) non-participants in the same period.	<ul style="list-style-type: none"> + no selection bias caused by unobservable characteristics + Ashenfelter's dip is not a problem - needs data for participants and non-participants before and after the programme (e.g. labour-market histories) - unobservable characteristics might bias results
Before-after comparison	Compares the outcome for participants before and after the programme took place.	<ul style="list-style-type: none"> + easy to implement + low data requirements (no information for non-participants needed) - general economic changes before/ until/after the programme period might be falsely attributed to the programme - Ashenfelter's dip (change of behaviour in the period preceding the programme)
Difference – in-differences	Compares the change of before-after outcomes for participants with the before-after change of outcomes for non-participants.	<ul style="list-style-type: none"> + takes account of selection on unobservable characteristics - needs data for participants and non-participants before and after the programme took place - Ashenfelter's dip
Cross-section	Compares the outcome for participants in the period after the programme with the outcome for non-participants in the same period.	<ul style="list-style-type: none"> + economy-wide changes are not attributed to the programme + Ashenfelter's dip is not a problem - needs data for participants and non-participants for the period after the programme

NB: For a discussion of 'Ashenfelter's dip' see Section 1.3.3.3, Figure 2.1. Source: Hujer et al., 2004a: Section 2.2.7.

to participate in a programme is, in most countries, beyond control of the evaluator.

The essential feature of social experiments is that evaluators deliberately control and manipulate the target populations as well as the conditions which determine the effects of an intervention. Out of the total group eligible for an intervention, such as the unemployed undergoing a retraining measure, two sub-groups are formed: those people who will participate in the intervention and those who will not. Participants and non-participants are

selected randomly. If the size of the total group is large enough, randomisation will allow for an equivalent composition of the sub-groups. 'Equivalence' refers to people's observable and non-observable characteristics that are of relevance for the intervention, such as age, sex, educational and social background, motivation, abilities, etc.

By comparing both sub-groups with each other, changes in the treatment group associated with the intervention, such as improvement of earnings, employment

situation, competences, can be attributed solely to the intervention; here this is retraining. Problems related to internal validity, selectivity, unobservables, etc., are controlled for (eliminated) in the experimental design.

However, there are several reasons which restrict the use of experiments. First of all, there are ethical, moral and also legal scruples. In Europe, contrary to the US, these scruples lead to avoidance of social experiments. It is considered that people who are eligible to participate in a particular social programme cannot be excluded *ex ante* and that participation cannot be restricted to selected people just for the sake of applying valid evaluation methods. Second, interactions between the treatment group and the control group are possible; and one cannot sufficiently control for the effects if some people leave the treatment group and others, from the control group, enter the programme. Finally, there may be serious problems concerning the external validity of experiments (Section 1.3.3.1) which are due to the high complexity of social phenomena.

Because of these restrictions, researchers usually employ quasi-experimental designs. Experimental and control groups are not constituted randomly and other methods will have to be applied to evaluate the intervention effects, such as the construction of control groups and matching approaches (see below).

In contrast to the US, social experiments in European countries are limited to some pilot projects or specific interventions. Some social experiments have been carried out in labour-market programmes. Examples are: the Restart experiment for unemployed people in Great Britain (cf. for example White and Lakey, 1992; Dolton and O'Neill, 1996); social experiments on the evaluation of training measures in Norway (cf. Torp et al., 1993); and a limited labour-market policy experiment in Sweden (Björklund and Regner, 1996). In addition, a number of model projects (*Modellversuche*) have been conducted for some time, particularly in Germany. They are

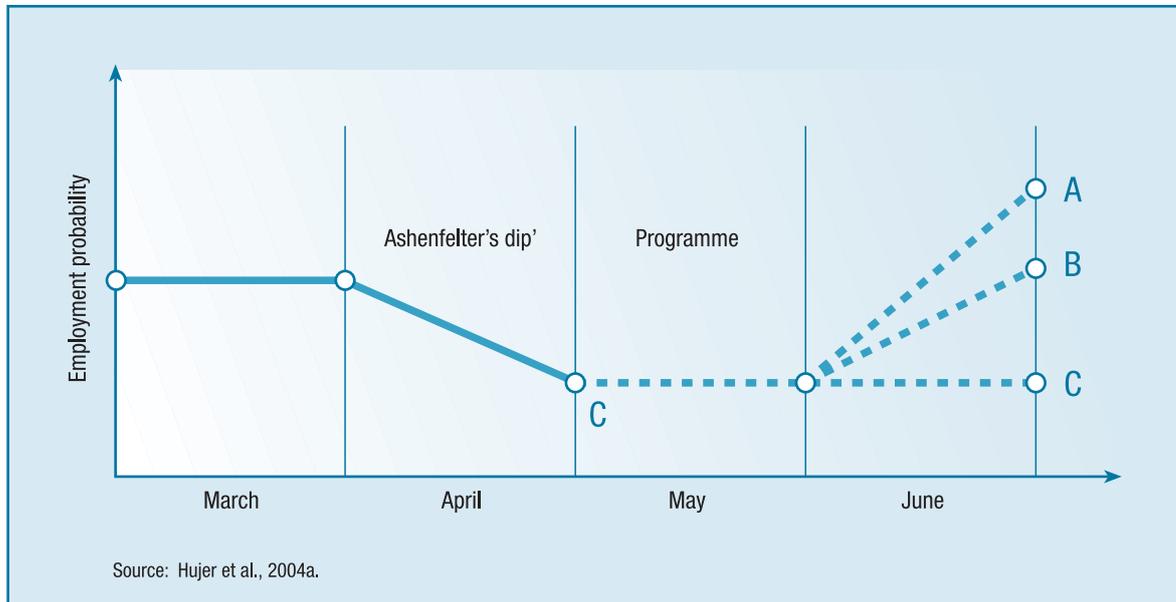
characterised by controlled conditions and processes and by continuous monitoring. They display some similarities with social experiments.

1.3.3.2. *Quasi-experimental designs: the matching approach*

When the matching approach is used to create experimental and control groups, designs are called 'quasi-experimental' approach. The basic idea is to identify within non-participants those individuals who share with participants all relevant characteristics (statistical twins) before the programme started. These 'twins' are not necessarily real persons but clusters of characteristics for participants and non-participants. As a result, differences in outcomes between participants and comparable control group members can be attributed solely to the programme. However, if the assignment of participants and non-participants is not random, the outcomes of non-participants cannot serve as counterfactual to those of participants. The matching approach, therefore, assigns covariates⁽³⁴⁾ randomly to participants and non-participants. Covariates in our case are, for example, pre-programme characteristics of individuals such as qualification, age, social background, etc., which may also influence the outcome and therefore have to be controlled for.

The role of matching is now to balance the distribution of all 'relevant' characteristics in both groups of 'statistical twins'. 'Relevant' characteristics are all those that influence either the assignment of persons to a programme or his/her potential outcomes. They may include characteristics from the pre-programme period, such as the previous employment status. For participants and non-participants the distributions of the characteristics are the same, i.e. they are balanced across the groups. In consequence, independence between potential outcomes and assignment to a group is achieved, resulting in an unbiased estimate.

⁽³⁴⁾ Covariates are other independent variables, e.g. individual characteristics, in a regression approach which may (partially) explain the dependent variable (e.g. outcomes).

Figure 2.1. **Ashenfelter's dip**

Hujer et al. (2004a: Section 2.2.4) conclude: 'The matching estimator is very data-demanding in the sense that we need information for participants and non-participants before and after the programme took place. In the case of exact matching, a "rich" data set is needed to ensure that we find comparable individuals in the control group for every combination of observable characteristics. [...] Problems arise, if a programme is compulsory, for example if all the unemployed have to participate in a training programme at a certain point in their unemployment spell. In this case it might get difficult to find a suitable control group, because all the unemployed will be in the programme at some time.'

1.3.3.3. 'Before-after' comparison

The comparison of programme participant outcomes before and after an intervention is widely used in evaluation because it is the most straightforward approach. The situation, in terms of earnings, economic and social position, etc., in the period before an intervention is taken as a point of reference for comparison with the outcomes after an

intervention. The effects – change of earnings, etc. – are called 'treatment effects on the treated'. Other related methods are:

- retrospective before-after comparisons where participants report on their situation before the intervention;
- the construction of panel surveys and time series that allow observation of pre- and after-programme outcomes, with several points of time where data are collected.

A major advantage of the before-after approach is that it does not require information on non-participants. However, longitudinal data, or at least data for identical individuals at two reference points of time, are needed to evaluate the outcomes for participants before and after participation in a programme ⁽³⁵⁾.

This approach relies on several implicit assumptions. One assumption is that the outcome categories (e.g. earnings, employment) before and after the intervention are not changed by external factors of influence, such as the overall economic situation.

Another assumption is that the outcome category, such as earnings or employment opportunities, is not affected by the

⁽³⁵⁾ The before-after method might also work with repeated cross-sectional data from the same population, not necessarily containing information on the same individuals (see Hujer et al., 2004a, for references).

anticipation of participation in the programme. This may, however, be the case. For example, individuals may behave in a different way to become eligible for the programme or change their behaviour (e.g. job search activities) in anticipation of their future participation.

This problem is well known as ‘Ashenfelter’s dip’ (Hujer et al., 2004a: Section 2.2.2). Figure 2.1 illustrates this dip by way of the hypothetical example of a programme which aims to enhance employment. The programme takes place in May. Shortly before start of the programme, in April, the employment situation of individuals who will participate deteriorates, as they reduce their job search activities in order to become eligible for the programme. If the dip is transitory and only experienced by participants, it will be misleading to take the employment situation in April as a basis for comparison with the effect after the programme.

After the programme, in June, the programme effect (employment probability) might be positive as indicated in case A: it is higher than in the period before the dip took place. In case B, there are no programme effects at all as the employment probability is the same as before the dip, in March. In this case, a comparison of the periods ‘before-after’ the programme, i.e. April and June, would overestimate the true programme effect. Case C indicates a deterioration because the employment probability is the same as after the dip, i.e. end April.

The problem of Ashenfelter’s dip can be avoided, if a time period before the dip took place is chosen as a reference level (here the situation in March).

1.3.3.4. *The difference-in-differences method*

An extension of the before-after approach is the difference-in-differences method. The before-after estimation may be affected by a selection bias: there might be unobservable characteristics of individuals which cannot be controlled for by the researcher. Examples are different expectations of the benefits of a programme which influence individuals’ decision to participate, or differences in their motivation to participate

The difference-in-differences method has been proposed to overcome selection bias arising from these unobservable characteristics. As with the before-after approach, the difference-in-differences method calculates the differences between outcomes of participants before and after the programme. These differences are compared with corresponding differences of non-participants (control group) in the same period. In its simplest application, average changes of an outcome category (e.g. earnings) for participants are contrasted with corresponding average changes of the same category for non-participants. This method is based on the assumption that the biases between participants and non-participants are, on average, the same in the periods before and after participation. By differencing the differences between both groups, these biases will be eliminated.

1.3.3.5. *Cross section method*

The basic idea of the cross-section method is to compare the average outcomes for participants’ (e.g. earnings) after the programme took place with those of non-participants in the same period. This means that the observed average earnings of non-participants represent the counterfactual outcome. This is useful if no information on participants in the preprogramme period is available or if macroeconomic conditions change substantially over time.

It is worth noting that by taking into account observed differences in individuals’ characteristics, such as different skill levels, systematic biases in outcomes may be eliminated. The assumption, again, is that no unobservable factors such as motivation or expectation of future benefits should lead people to participate in a programme. If this were the case, we might even, in the absence of any treatment effect, observe a higher average outcome of participants compared with non-participants (selection bias). Ashenfelter’s dip (see above) is not problematic in this approach as participants and non-participants are compared after the programme took place. Moreover, if external

influences (e.g. overall economic situation) affect both groups equally, the method is not vulnerable to problems that affect the before-after method.

1.3.4. Evaluation at aggregate level ⁽³⁶⁾

Evaluation of programme or policy impact at an aggregate level ⁽³⁷⁾ complements evaluation at micro/individual level in that it also takes into account positive or negative indirect effects on non-participants, the so-called ‘spill-over effects’. The question under investigation is whether or not a positive effect found at micro level, for participants, is also positive at aggregate level for the whole population or workforce; in other words, what is the net effect of a programme or policy? ‘Aggregate’ here refers to overall impacts of a programme for a country or region; in this sense it is a ‘macro perspective’. Once all programme effects at micro and aggregate level have been analysed, the efficient allocation of resources, taking into account all benefits and costs, may be subject to a cost benefit analysis (CBA). For an overview on different effects occurring at aggregate level see Box 2.3.

1.3.4.1. Indirect effects

Whereas positive effects on participants are a strong justification for the programme conducted, one should consider whether this programme also generates negative indirect effects, for example for non-participants. Indirect effects are effects which were not explicitly intended by the programme and which might question the costs and efforts devoted to an intervention. They have similarities with external effects or spillovers in macroeconomic growth models (see Section 2.1.1).

Taking the example of programmes within active labour-market policy (ALMP), which usually also contains specific measures of training or skill enhancement, we can distinguish between several negative indirect effects: dead-weight, substitution, displace-

ment and tax effects (Calmfors, 1994).

A dead-weight effect occurs when programme participants would have been recruited also without the programme. If the improvement in employment prospects was an objective of the programme, the programme had no such effect, at least for the respective participants.

A substitution effect may occur, for example, in a situation where a programme subsidises jobs in order to enhance the employment prospects of the target group. If an employer recruits a programme participant instead of an unsubsidised worker who would have been hired otherwise, substitution effects are at stake. The same would be the case if an employer dismisses workers in favour of subsidised workers. The substitution effect refers to the ‘extent to which jobs created for a certain category of workers simply replace jobs for other categories of workers, because relative wage costs are changed’ (Calmfors, 1994).

The displacement effect refers to displacement in the product market where companies with subsidised workers – and resulting lower labour costs and/or higher productivity – are able to increase their output at the expense of companies which do not employ subsidised workers ⁽³⁸⁾.

Finally, a taxation effect affects non-participants when they pay taxes to finance a programme but receive no treatment.

The appropriate method for investigating these indirect effects at aggregate level is dynamic panel surveys among all workers – participants and non-participants – over time. This requires aggregate data for the variables of interest. For example, aggregated flow data on transitions from and into employment and unemployment are of particular interest in evaluating ALMP measures.

These kinds of data for all workers are able to reveal information on net aggregate effects for a country or region after the programme took place. They do not, however, indicate

⁽³⁶⁾ This section mainly draws on Hujer et al. (2004a: Sections 2.3 and 2.4).

⁽³⁷⁾ Not to be confused with impact research at macroeconomic level which does not focus on specific programmes or interventions but at overall effects of, for example, education and training or R&D on economic growth; see Chapter 2.

⁽³⁸⁾ This is similar to the crowding-out effect known in macroeconomics.

causal relationships on, for example, the concrete replacement of a non-participant by a programme participant in a particular company. Such information can only be gathered at micro level, such as job biography or company surveys.

1.3.4.2. Cost-benefit analysis

Microeconomic evaluation intends to uncover the outcomes of programmes for the individual. In addition, aggregate evaluation studies can reveal positive or negative indirect effects on the whole economy or region. These are necessary and complementary steps in evaluating the merits and values of an intervention. However, they do not include costs; for this, a cost-benefit analysis (CBA) is required.

Cost-benefit analysis ‘widens the perspective of an impact analysis. A CBA is a method that provides a consistent, explicit and transparent procedure to evaluate [...] projects in terms of their consequences, i.e. in terms of their costs and benefits.’ (Hujer et al., 2004a: Section 2.4) CBA intends to assess the net present value of all benefits less all costs, valued by a single monetary measure. A CBA for social policies (societal bill) also includes wider benefits such as equity and distributional goals and thus considers all societal costs and benefits.

Although it is a useful tool for assessing social programmes, the application of CBA is, in practice, rather limited, first, because of its methodological complexity and, second, because CBA requires thinking in alternatives. Its acceptance by decision-makers, therefore, heavily depends on the institutional and political situation.

A CBA includes a number of features and stages (Hujer et al., 2004a: Section 2.4).

Timing is important. As with general evaluation (Part 1, Section 1.1), a CBA can be conducted *ex ante*, *ex post* and during a programme. An *ex ante* CBA intends to justify the implementation of a particular intervention. However, because most benefits and costs will arise in the future, there are considerable uncertainties of estimation. An *ex post* CBA is conducted when all costs are irreversibly sunk.

Box 2.3. Programme effects at aggregate level

The following effects are distinguished in an evaluation:

Real and monetary effects: real effects are impacts on the social welfare, i.e. the final utility gain or loss of a programme; monetary effects only change the relative prices but have no welfare effects.

Direct and indirect effects: direct effects refer to the explicit objectives of a programme; indirect effects cover non-intended impact of programmes. Examples for indirect effects in the context of ALMP are: deadweight, substitution and displacement effects (see above).

Tangible or intangible effects: tangible effects refer to physical or measurable impacts such as earnings or employment probabilities; intangible effects include impacts on health, crime, trust, social cohesion, political participation, etc. (Section 2.1).

Final or intermediate effects: final effects occur at the level of the consumer; intermediate effects occur at the producer level.

Internal or external effects: internal effects concern impacts within a prespecified target group; external effects are positive or negative impacts outside the target group or the subject of a programme, such as crime reduction or better parenting.

Source: Hujer et al. (2004a: Section 2.4).

However, this suffers similar uncertainties as *ex ante* CBA because medium and long-term benefits are not yet known and have to be forecast. CBA during the intervention serves to improve particular features of the programme or even to cancel the programme.

The selection of alternatives is also a major factor. CBA is only of use if the alternatives to a particular intervention have also been subject to a (mostly *ex ante*) CBA. Otherwise there would be no point of reference to assess and compare the results of a CBA. However, the decision which programme to carry out is thus restricted only to these defined alternative measures (although there might be others and more appropriate alternatives *ex post*).

The definition of the reference group is the basis for deciding whose costs and benefits should be considered. These could be the participants, non-participants or the society as a whole. The latter two references require an

aggregate CBA at macro or regional level taking into account all indirect effects mentioned in the above sections. A social CBA looks not only at private costs and benefits for individuals or firms, but includes all programme impacts whether they are private or social, tangible or intangible, direct or indirect, intended or not.

An important step in a CBA is to define all the effects of a programme in terms of costs and benefits (Box 2.3), and to forecast these over the lifetime of the project and perhaps beyond. For some impacts, a clear cause-effect relationship may exist, for others such as intangibles (e.g. crime, health, etc.) an assessment of benefits is much more difficult (Section 2.1.2).

The most crucial step in a CBA is to measure and aggregate individual costs and benefits. All these impacts have to be measured in monetary terms. Tangible effects can be measured in market prices, external effects can be proxied by shadow prices. For example, the effects of declining delinquency in a region could be approximated by the increase in real estate prices after a programme, if surveys have shown that real estate purchases also depend on the criminality rate in that region. The application of market or shadow prices is, however, difficult for intangible effects. One method would be to ask people directly how much they were willing to pay for a certain improvement, such as political stability. Another possibility is the observation of preferences. If, for example, in a past programme EUR 1 000 000 have been spent to increase membership in associations (as an indicator of social integration), and 10 000 people became members, one could infer that the value of social integration is EUR 100 per capita. However, such indirect measures have to be made with care.

Another problem associated with the measurement of individual costs and benefits is their aggregation. This presupposes

weighting each individual, normally done equally. If the net effect of a programme is positive, the sum of all individual benefits exceeds the sum of all individual costs. In this case, because the overall benefits increase, losers from the programme could theoretically be compensated by winners, for example via tax redistribution. If problems of measuring these costs and benefits cannot be resolved, a cost-effectiveness analysis can be carried out where the benefits of an intervention are not monetarised. An example would be to calculate the costs of a programme per person employed afterwards;

When comparing costs and benefits, the problem occurs that both categories do not accrue at one point in time but in a period or even in different periods. In order to compare future costs and benefits with those which occur in the present, the flows of costs and benefits can be discounted, by taking a discount rate (r) such as the market interest rate⁽³⁹⁾. The discounted flow of costs (C_t) and benefits (B_t) at different points of time (t) represent the net present value (NPV) of a programme:

$$NPV = \sum_t \frac{B_t - C_t}{1 + r}$$

If the net present value is positive, the internal rate of return exceeds market interest rates and may indicate a favourable investment compared with alternative investments, for example in physical capital.

Because of the many caveats, assumptions and uncertain predictions underlying a CBA, sensitivity analyses can be carried out to demonstrate how sensitive the NPV results are to different assumptions about key parameters. Careful scenarios, which vary the most important assumptions, are one way to 'protect' the results against doubts.

⁽³⁹⁾ However, market interest rates do not necessarily correspond to rates of return because of imperfect capital markets. Moreover, whereas costs occur during programme implementation, most benefits are spread over the future. Small changes in the assumed discount rate, therefore, might have a great impact on the net present value of a programme.

1.3.5. Choice of methods and data requirements ⁽⁴⁰⁾

1.3.5.1. *Choice of method*

When applying quantitative methods, the main issues of concern are to address the basic evaluation problems discussed in Section 1.3.1 and the other various biases and sources of errors discussed throughout this chapter.

Since we cannot observe the same individual simultaneously both with and without participation in a programme, the 'true' treatment effect has to be calculated. Of the methods presented above, experience has shown that the matching approach appears to be the most reliable ⁽⁴¹⁾. This approach avoids problems caused by unobserved characteristics and thus a potential selection bias. However, its application requires a rich and informative data set which contains all variables and characteristics expected to be relevant both for the decision on participation and the outcomes. Furthermore, if unobservable characteristics are assumed to cause a selection bias, an additional difference-in-differences procedure might be useful.

The various problems of heterogeneity can be tackled by distinguishing as far as possible between programme outcomes of different groups, contexts and regions.

A general recommendation, whatever the method chosen, is that the more informative the data set, the more likely biases will be avoided and internal and external validity be obtained.

Although quantitative approaches are suited to uncovering the 'true' effects of an intervention, policymakers and stakeholders who expect to receive an overall 'simple' statement on whether a programme has worked or not, might be dissatisfied with rather complex conclusions.

1.3.5.2. *Data requirements*

A crucial issue when using quantitative methods is the quality and coverage of data used in an evaluation study. Especially when

applying the matching procedure and taking into account the heterogeneity problem, an informative data set is needed. Both the treatment and control group must contain a sufficient number of individuals; there is no rule of thumb on how many control group members are needed. Clearly, the more comparable both groups are in terms of their relevant individual characteristics, the more likely are 'good' matches.

For example, if the programme addresses the long-term unemployed and the control group contains only the long-term unemployed as well, the control group does not need to be much larger than the treatment group. If, however, the control group is selected randomly from the whole population – including the unemployed as well as employed and non-employed people – good matches are less likely and the sample size of the control group has to be considerably bigger.

A well-known problem is that evaluators are not normally involved in the creation of data sets to be used later for evaluation. This refers to data on the characteristics of both participants and control group members. Evaluators can either rely on existing data sets or build up new ones:

- (a) existing data sets, such as surveys or panels, do not always contain the specific characteristics needed for evaluation. Moreover, the number of observations may be too small to, for example, compare people (statistical twins) with specific characteristics in a matching approach;
- (b) building up new data sets has, of course, numerous advantages. The evaluator gets full control over the information to be collected for the treatment and control groups, including all relevant and comparable observable characteristics. However, the major and crucial disadvantage that impedes this attempt is the high cost.

Finally, the outcome variable has to be reliable and should be observed for a sufficient

⁽⁴⁰⁾ This section draws mainly on Hujer et al. (2004a, Chapter 4).

⁽⁴¹⁾ Leaving aside true experiments for the ethical reasons mentioned above.

period of time after the programme in order to judge sustainability. It might be advisable that conclusions distinguish between short-, medium- and longer-term impacts.

1.4. Discussion

To conclude this chapter, it is useful to discuss the use of qualitative and quantitative methods according to the phases of VET programmes. Our aim is to demonstrate the complementary nature of both approaches for gaining a full understanding of the programme under evaluation in formative and summative perspectives.

1.4.1. Logic and complementarity of qualitative and quantitative methods

We have discussed features, design and particularities of qualitative and quantitative approaches to evaluation throughout this chapter. However, quantitative and qualitative approaches are not mutually exclusive but alternative and complementary strategies for research and evaluation. In this respect, their fundamental differences in terms of logic and purposes help explain better how they can complement each other.

While quantitative methods are used to generalise and compare results in a standardised way, the purpose of qualitative approaches is to increase the depth of understanding. Quantitative methods usually require large samples, selected randomly, to generalise with confidence the results from the sample to the respective part of the population⁽⁴²⁾. In contrast, qualitative approaches mainly use small samples, sometimes even single cases, which constitute rich-information cases, selected purposefully to permit in depth enquiry into and understanding of a phenomenon (Patton, 2002: 45 f.).

Quantitative methods tend to favour controlled settings, holding constant or eliminating other external influences, as well as a limited set of variables (predefined as dependent or independent) and outcomes to be measured⁽⁴³⁾. In a qualitative evaluation design, the evaluator does not want to control *a priori* the process under analysis. He/she wants to remain open to all kinds of outcomes and does not impose constraints on the number of variables that will be analysed (Patton, 2002: 39-45). Furthermore, context-sensitivity is an important feature of qualitative approaches.

The logic of qualitative enquiry is more one of discovery, trying to uncover processes and particular dynamics of cases, being open to realities as they appear to the participants or other stakeholders. Quantitative methods adopt more a logic of verification and research to discover regularities or even 'laws' or to make generalisations, trying to understand what works and what does not work and thus to ensure accountability⁽⁴⁴⁾. However, both work towards designing future interventions and improving current ones.

There are mutual critiques and misunderstandings between the respective supporters of quantitative and qualitative methods. The primary critiques of quantitative evaluations are: oversimplifying the complexities of real-world programmes and participants' experiences; missing major factors of importance that are not easily quantified or observable; and failing to portray the programme and its impact as a whole.

In contrast, qualitative inquiry is taxed of being subjective and unscientific; qualitative 'soft' data can be dismissed as mere anecdote. Evaluators and researchers can be locked in their philosophies, paradigms, schools of research and preferences. However, when evaluating programmes and policies, pragmatism is required.

⁽⁴²⁾ These are characteristics of large programmes or macroeconomic quantitative methods. However, quantitative methods can be also applied to small programmes when evaluating e.g. labour-market entries of participants, without generalising this to the whole population.

⁽⁴³⁾ See Table 2.1 where various types of control are presented.

⁽⁴⁴⁾ We are not going to discuss here the paradigms and philosophical assumptions underlying each approach. This is done in Part 1 of this report.

1.4.2. Evaluation process and basic methodologies

In Part 1, Section 1.1, we addressed the main phases of the process accompanying the implementation of VET programmes. Each phase implies different evaluation methodologies ⁽⁴⁵⁾.

The first phase of formulation, planning and preparation of an intervention requires conceptualisation, setting objectives and programme design. At this stage, preformative evaluation or *ex ante* evaluation using mainly qualitative methods has to investigate whether the input conditions such as human and financial resources, legal frame, theoretical basis, etc., are suited to programme objectives. Preferably at this stage the ground should be laid for summative evaluation methods. This could include the constitution of control groups and the preparation of data collection which allows the impact of the intervention to be judged.

During implementation and execution, evaluation has to assess whether the programme design, organisation and resources are appropriate; whether the target group has been reached; whether the objectives set in the planning phase are suited to ensure success, and whether the expected learning process and behaviour change is taking place. This entails qualitative – but where appropriate also quantitative – feedback from stakeholders, programme administrators and staff, including participants. Formative evaluation is used here with the purpose of improving and, if necessary, redirecting or redefining the programme implementation parameters. Furthermore, summative evaluation of preliminary results can be carried out. Evaluation during a programme serves to improve its execution or to decide whether or not it should be continued.

After completion of the programme, analysis of its impact is subject to summative evaluation using social science research methods. These include various methods to judge the effects on participants compared with the counterfactual situation (e.g. with non-

participants or previous situation) and the investigation of unintended effects. Methods are primarily quantitative, although some more qualitative methods may also be used, such as satisfaction surveys, self-assessments, Delphi inquiries. The results and experiences gathered can be the basis for decisions on the follow-up of the intervention or can be transferred to other similar social programmes. In this case, evaluation takes on a formative character.

An essential requirement for scientific evaluation is that the choice of both explanatory and outcome variables should be based on theoretical considerations. Which individual characteristics are expected to influence participation decisions and outcomes, and why? And which kind of outcomes (e.g. earnings, employment probability, learning success, etc.) are at the centre of evaluation? Theoretical considerations can be derived from the relevant domains of social science theories such as human capital or segmentation theories; the relevant variables can then be identified and the interaction between different variables, as well as with the social and economic framework, explained.

‘An evaluator needs a large repertoire of research methods and techniques to use on a variety of problems. Thus an evaluator may be called on to use any and all social science research methods, including analysis of quantitative data, questionnaires, secondary data analysis, cost-benefit and cost-effectiveness analysis, standardised tests, experimental designs, participants observations, and in-depth interviewing. [...] Such an evaluator is committed to research designs that are relevant, meaningful, understandable, and able to produce useful results that are valid, reliable, and believable.’ (Patton, 2002: 68) ‘Methodological appropriateness [should be] the primary criterion for judging methodological quality, recognising that different methods are appropriate to different situations.’ (ibid.: 72).

⁽⁴⁵⁾ See Section 1.1 and Table 1.1 in Part 1 that summarise these processes and the associated evaluation types and methodologies.

Concluding, evaluators might balance between extreme positions: on the one end, rather rigorous scientific methods, such as social experiments; on the other end, pragmatic qualitative approaches reflecting alternative forms of action. This choice does not have to be made. Evaluation can meet both strict scientific requirements and be of maximum worth for the commissioner and other interest groups. Today, most researchers agree that there is no rigorously set methodology. 'A good evaluation will combine (quantitative and qualitative) approaches to arrive at a set of conclusions that take account of context, implementation, process and short- and long-term outcomes.' (Plewis and Preston, 2001: 27). In a similar way, Chelimsky (1995: 6, cited in Stockmann, 2000b: 19) points out: 'we think less today about the absolute merits of one method versus another, and more about whether and how using them in concert could result in more conclusive findings'. Instead of methods and techniques, the proper evaluation issues are to be placed in the foreground: 'we have learned that the choice of methods (and measures, instruments and data) depends much more on the type of question being asked than on the qualities of any particular method' (Mertens, 1998: 12, cited in Stockmann, 2000b: 13).

2. Impact research: approaches and methods

There is no common definition of impact research ⁽⁴⁶⁾. For this report we have adopted a definition that is distinct from evaluation research insofar as impact research is not focused on specific programmes or interventions but investigates to the impact of education, training and skills on economic and social development, on company or individual performance. The borderline between impact and evaluation research is not always clear cut. Thus, for example, provision of training by

a company may be part of a specific programme of this company. Our approach is to discuss a broader range of impacts of education, training and skills, i.e. of human capital, irrespective of whether they are part of specific programmes or interventions.

This chapter discusses the most common approaches and methods used in impact research to measure the impact of education, training and skills on society, companies and individuals. This includes discussion of the material as well as non-material benefits of education and training. Figure 2.2 summarises the various returns on education, training and skills at micro and macro level; results will be presented in Part 4.

2.1. Social and economic benefits of education and training: basic approaches

The contribution of education, training and skills to economic prosperity and social inclusion are, and have always been, one of the main motives for public and private investment in education and training by the individual, enterprises and the State.

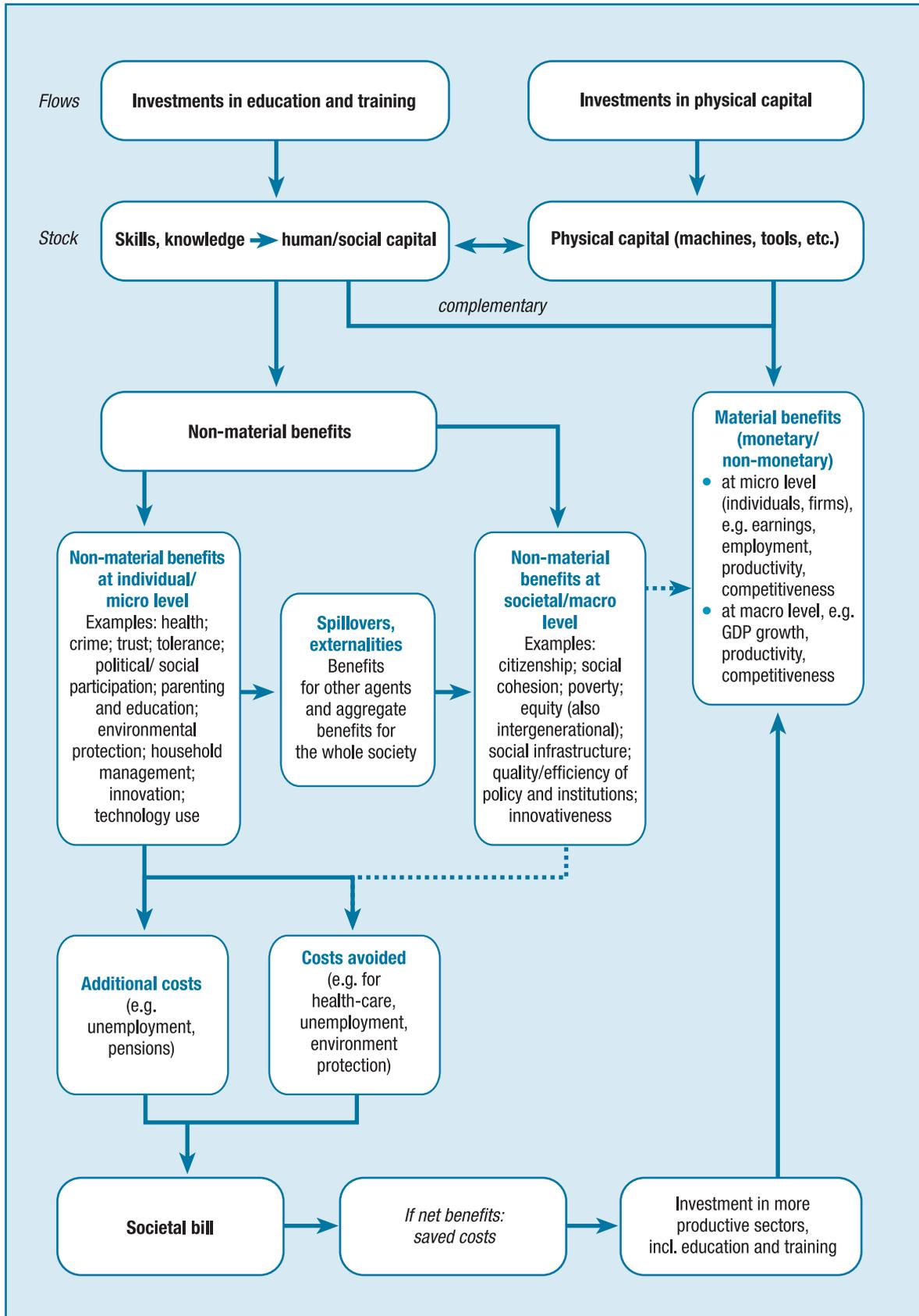
At micro level, returns on education and training are mainly measured by an increase in earnings or wages. However, analyses at individual level at most cover private returns on education and tend to underestimate the full returns for society, including indirect effects (social returns). In this context we should mention that – as Wolf (2002) rightly points out – rates of return ⁽⁴⁷⁾ on education and training are a poor, and even inappropriate, concept for measuring the contribution of education and training to economic growth. Statements on the significance of education and training for economic growth should, therefore, be based on macroeconomic growth theories and not on microeconomic rates of return analyses.

At macro level, the contribution of education, training, knowledge and skills to

⁽⁴⁶⁾ 'Impact research' is used for a variety of activities in technology, psychiatry, environment, learning, etc. Activities range from measuring the impact of specific measures to the impact of global changes of phenomena, e.g. climate, environment, etc.

⁽⁴⁷⁾ See Section 2.3.1 for a discussion on methods to calculate the returns on education and training.

Figure 2.2. **Benefits of education and training**



economic growth, productivity and development are in the foreground. Estimates look at the contribution of all production factors – including human capital – to the growth in production of goods and services (GDP or GNP; Box 2.4) in the economy. They also (should) take into account external effects: education and training acquired by individuals might also be of benefit (spill over) to other ‘agents’ in a society (individuals, companies, etc.) than those who have acquired higher skills and qualifications. These spillovers or externalities at macro level – if positive – provide additional justification for public support for education and training (Sianesi and van Reenen, 2002).

Box 2.4. GDP and GNP as indicators for economic growth

Normally, growth analyses refer to the level and/or to changes in GDP (Gross Domestic Product), sometimes also to GNP (Gross National Product). GDP represents the total value of goods and services produced within a nation. GNP represents the total value of all goods and services produced by a nation’s economy. The difference between GDP and GNP is the net transfer of incomes to and from other countries. GDP includes income transfers from abroad minus income transfers to abroad. GNP includes income transfers to abroad and excludes incomes transferred from abroad. ‘Incomes’ denote the incomes of all production factors: labour, physical capital and land (i.e. earnings, profits, rents, etc.).

In order to compare economies of different sizes, GDP or GNP are often related to the workforce: ‘GDP or GNP per capita’. This ratio represents productivity at macro level and denotes the value of goods and services produced by a worker in a given period (output per worker). This ratio is therefore influenced by the total GDP, or GNP, and its change as well as by a change of the number of people employed.

2.1.1. Modelling the determinants of economic growth and productivity

Linking education, training, knowledge and skills – summarised as ‘human capital’ (Box 2.5) – and overall economic performance of an economy is the domain of modern macro-economic growth theories. Following Sianesi and van Reenen (2002) we have to distinguish

Box 2.5. Defining human capital

Definitions of human capital range from more economic-oriented ones to those encapsulating diverse characteristics of human beings. Some examples found in www.google.com/search are:

- stock of accumulated skills and experience that make workers more productive;
- expertise, experience, capability, capacity, creativity, adaptability, etc., possessed by individuals in an organisation;
- individual or collective knowledge, skills, physical attributes of people used in producing goods and services;
- quality of labour resources which can be improved through investments in education, training and health;
- time experience, knowledge and abilities of an individual household or a generation which can be used in the production process.

In 1998, the OECD (1998: 9) adopted the following definition of human capital: ‘the knowledge, skills, competencies and other attributes that are relevant to economic activity’. This definition refers not only to the level of education of a person, but also to his/her capacity to put a wider range of skills to productive use (Healy, 2000: 19). The OECD recently developed a wider definition that also refers to the personal benefits (OECD, 2001d: 18): human capital means ‘the knowledge, skills, competencies and attributes embodied in individuals that facilitate the creation of personal, social and economic well-being.’

between (a) the neo-classical approach formulated by Solow (1957) and its augmented form, and (b) the endogenous growth theories, developed since the late 1980s. Neo-classical models imply that ‘a one-off increase in the stock of human capital leads to an associated one-off increase in productivity growth, whereas endogenous growth models suggest that the same one-off increase in human capital can lead to a permanent increase in the productivity growth rate. In the short-term, both model types produce broadly similar results, each dependent on the precise model specification, but, over the longer-term, the newer growth models imply significantly higher returns to investments in human capital’ (Wilson and Briscoe, 2004: Section 6.1). This is ‘based on the notion that a higher rate of innovation is associated with economies richer in human

capital, increasing level of human capital is expected to have an effect on the growth rate of productivity' (Sianesi and van Reenen, 2002: 5).

2.1.1.1.1. *The neo-classical model: growth accounting*

The neoclassical model (Box 2.6) decomposes the growth of national output (GDP or a related measure) into its constituent parts, by weighting the increase in each input factor with its relative factor share ⁽⁴⁸⁾. Input factors are physical capital, labour and a 'residual factor' including all other factors of influence.

Following Sianesi and van Reenen (2002: 7), such accounting exercises based on neoclassical growth theory and using a production function, aim to assess the relative contribution of inputs (physical and human capital) versus the residual factor (total factor productivity, TFP) to either (a) growth in output (GDP) or (b) cross-country differences in output per worker.

Neoclassical models of economic growth, as developed by Solow and Swan, identified technological progress as a residual not explained by the production factors capital and labour. Though 'capital' theoretically includes both human and physical capital, human capital was not, in its augmented form, considered in many empirical studies based on this model.

Despite efforts to investigate further the 'residual factor' and to diminish its importance by extended frameworks – for example by including the quality of inputs and by investigating technical change – the issue remains that 'accounting is no explanation' (Griliches, 1997a). Growth accounting exercises provide no information on the potential indirect effects that education can have on the level or growth of GDP, for example through labour force participation, R&D, etc., 'the existence of [...] a positive correlation tells us nothing about causal relationships, about the mechanisms, the processes through which human capital accumulation affects economic growth.' (Sianesi and van Reenen, 2002: 8).

Box 2.6. **The neo-classical model: growth accounting**

The original Solow model distinguishes between three input factors: the aggregate stock of physical capital K , the labour force L , and the factor A which captures 'technical progress' and other omitted variables. A is an index of technical efficiency or 'total factor productivity' (TFP) and summarises the current state of world technology (world technology frontier). A also includes possibly omitted variables such as geographical location, climate, institutions and natural resources (de la Fuente and Ciccone, 2002: 27). Because A itself is not modelled, this factor is called the 'residual factor', responsible for all variations in GDP which are not explained by physical capital and labour.

The determinants of GDP growth (Y) are put together in a Cobb-Douglas production function of the form

$$Y_t = K_t^a L_t^b A_t$$

where the coefficients a and b measure the elasticity of GDP growth (output elasticity) related to a change of the respective input factor. For example, an increase of 1 % in the stock of physical capital (K) would increase GDP (Y) by a %, holding constant the labour force (L) and the level of technical efficiency (A). For estimation purposes this production function is calculated by growth rates or logarithms.

In its augmented form the neoclassical model extends the basic production function by an additional input factor: the stock of human capital (H), mainly measured by years of schooling. However, this estimation does not take into account human capital externalities which increase the level (and not merely the growth) of output. The production function in its augmented form is

$$Y_t = K_t^a L_t^b H_t^c A_t$$

where c measures, *ceteris paribus*, the GDP growth associated with an additional year of schooling.

Source: de la Fuente and Ciccone (2002).

2.1.1.1.2. *New theories on endogenous growth*
Macroeconomic studies, in particular regressions analysing the impact of human capital, intend to assess the impact of externalities and wider benefits on economic

⁽⁴⁸⁾ The following discussion draws mainly on Sianesi and van Reenen (2002: 5 ff.); de la Fuente and Ciccone (2002); Wilson and Briscoe (2004).

growth ⁽⁴⁹⁾. To capture these benefits, macro growth regressions are calculated instead of merely accounting the contribution of input factors as done in the framework of neoclassical growth theories. The endogenous approach seeks to verify the additional impact of human capital on growth beyond the static effect on the level of output (GDP). It assumes that economies with high human capital are characterised by a higher rate of innovation and R&D; increasing the level of human capital is thus expected to have an effect on the growth rate of productivity or GDP.

Human capital plays a prominent role in the new growth models which account for technological progress. Romer (1990) recognised human capital as a primary source of technological progress and therefore economic growth. In particular, research workers and R&D are seen as the source of new ideas and hence profits.

In these approaches, GDP growth is determined endogenously, i.e. within the model, instead of being driven by residual, exogenous technological progress as assumed in neoclassical models. However, disagreement remains on the mechanisms by which human capital affects economic growth. There are two main schools in endogenous growth theory. One view (e.g. Lucas, 1988) sees the ‘accumulation’ of human capital as the source of economic growth, whereas other approaches (e.g. Nelson and Phelps, 1966; Benhabib and Spiegel, 1994) assume that it is the ‘stock’ of human capital that determines the ability of an economy to develop and assimilate technologies, so producing economic growth. In endogenous growth theory, education and training influence the economic growth of a country via two main channels:

(a) human capital is explicitly incorporated as an input factor by modelling individual investment choices measured, for example, by enrolment rates; this refers to the accumulation or flow of human capital. In addition, the external effects of human capital are often modelled explicitly. The

theoretical prediction that GDP growth is (also) a function of human capital thus is similar to neoclassical growth accounting exercises;

(b) the factors that influence endogenous growth, in particular technological change and productivity, are ‘explicitly related to the stock of human capital. This may be either because human capital is assumed directly to produce new knowledge/technology or because it is an essential input into a research sector which generates new knowledge/technology.’ (Sianesi and van Reenen, 2002: 8). Indicators of human capital stock are, for example, years of schooling or level of educational attainment of the population.

This distinction between accumulation and stock of human capital in these two major strands of endogenous growth approaches has important implications for policy. In the framework of human capital accumulation, a subsidy to education or an educational programme which raises the level of human capital will have a limited once-and-for-all effect on GDP growth. In the second approach, the growth rate of GDP will be increased forever, via dissemination of knowledge and technology. Thus, the first approach would imply policy actions to upgrade human capital in an economy, whereas the second approach would also require policies to foster R&D and technology.

However, Sianesi and van Reenen (2002) take a more sceptical view on the new growth theories. They suggest that most macro regressions only tend to enlarge the set of variables and do not try to test one theory against the other. ‘Most of these regressions include the stock of human capital as an explanatory factor and take inspiration and justification – albeit quite loosely – from the endogenous growth strand [...]. It is important to note that in such cases the estimated increase in productivity is not simply a phenomenon of the transitional period as the increase in the flow of education leads to a gradual increase in the equilibrium human

⁽⁴⁹⁾ See for the following discussion Sianesi and van Reenen (2002: 8 ff.).

capital stock. Implicitly it is claimed that increasing average education in an economy will permanently increase the rate of economic growth [...].’ (ibid.: 9)

Comparative empirical work tries to explain growth performances across countries. Unlike neoclassical models, which calculate the parameters (output elasticities) of a production function (Box 2.6), endogenous growth models try to explain the variation in total factor productivity across countries, which was unexplained in growth accounting approaches. Productivity is most often measured by GDP per capita. Variables in cross-country regressions include proxies of human capital (such as years of schooling), initial level of GDP, investment ratios in physical capital, geographical dummies and a number of variables concerning the role of governments, political stability, market distortions and economic systems. (Sianesi and van Reenen, 2002) The aim of such macro regressions is to shed ‘light on the origin of differences in growth rates across countries, and helping to identify those policy measures most likely to enhance growth’ (ibid.: 9).

Some recent studies exploit time-series for one or more countries. They can be used to explain both cross-country differences in growth and the evolution of economic performance over time in a single country.

2.1.1.3. *Methodological reservations*

A number of restrictions should be considered when calculating the contribution of human capital to growth and productivity. They can be addressed here only briefly ⁽⁵⁰⁾.

(a) First is measurement of human capital. In most cases, human capital is defined by proxies such as average years of education, enrolment rates or proportion of the labour force which has received primary, secondary or higher education. They focus on formal education and thus neglect wider definitions of human capital

including on-the-job training, experience and learning by doing ⁽⁵¹⁾. Equally, different types of education and quality of education are often disregarded.

- (b) Data quality, in particular when constructing time series and/or cross country comparisons for the numerous variables, is in general rather poor. Furthermore, few studies apply sensitivity tests in order to judge the robustness of their models to measurement errors. Finally, data are typically collected from a variety of sources which are not always consistent.
- (c) Endogeneity and simultaneity biases might lead to a reverse causality problem in that decisions to invest in education may (also) be influenced by the level and growth of income or GDP in a given country. In this case, the demand for education is a variable which depends on income, and not reversely, as suggested in growth models. Also, countries with higher income levels typically have service and modern high-tech production sectors which normally require better educated and trained workers. This structural change may require advanced education and training; thus, there may be an impact of economic growth on human capital accumulation and vice versa. It is most plausible that both influences are simultaneous and that there is a bidirectional causality between human capital accumulation and economic growth. Failure to control for this may cause a simultaneity bias. When longitudinal data sets are available, this bias can be reduced by using time lags of the endogenous variables, i.e. lags between an increase in human capital and subsequent GDP growth. However, these time lags may be long and unknown.
- (d) Parameter heterogeneity arises when cross-country growth studies include

⁽⁵⁰⁾ For more details see the literature indicated above, particularly: Sianesi and van Reenen (2002: 13 ff.); de la Fuente and Ciccone (2002: 25 ff.); Wilson and Briscoe (2004: Chapter 6).

⁽⁵¹⁾ In microeconomic wage equations, ‘work experience’ in most cases is proxied by calculating the difference between the actual age of an individual and the average/hypothetical age when leaving education. Sometimes, these estimations are modified by taking into account average periods of unemployment or non-employment for different qualification levels and sex.

countries at various levels of economic development and also education and training (in terms of content, sequence and quality). As results from most growth studies are an average from heterogeneous countries, the question is what such regressions tell us about the contribution of human capital to growth and of how reliable the results are for policies in a specific country.

- (e) There is no reason to assume a linear relationship between human capital and productivity or growth. In fact, some studies indicate that the contribution of a factor (physical capital, human capital) is likely to diminish over time and with increasing development. The impact may still be positive but the growth rate may decrease. This peak indicating decreasing effects of human capital on productivity is estimated for OECD countries to lie between 7.5 and 8.4 years (Krueger and Lindahl, 1998; Barro and Lee, 1994) ⁽⁵²⁾.

2.1.2. Addressing the wider social benefits and social capital

We have already mentioned, in the framework of macroeconomic benefits of education and training measures (Section 2.1.1), some issues of externalities and indirect effects of education and training. These indirect effects include benefits to people or firms other than those who undergo and/or invest in education and training. They include the wide range of 'non-material benefits' which accrue to society ⁽⁵³⁾ as the human capital of its citizens increases. Examples for these benefits are citizenship, trust in democracy and its institutions, political and social participation, social inclusion, etc. These benefits are objectives *per se* for society; they may also have impacts on economic growth and productivity.

2.1.2.1. Externalities, spillovers and social rates of return

The benefits of human capital are not necessarily restricted to individuals or firms investing in additional education and training but might yield benefits also to other workers, firms and even societies. For example, more educated or trained workers may increase the productivity of their less educated colleagues, a higher level of human capital in a firm thus entailing a higher incidence of learning from others. Or there may be spillover effects if technical progress or knowledge accumulation is caused by investments in human capital and spreads out to others.

Human capital investments may also have external effects on other people outside the work environment, most of which are non-monetary. More education can be associated with better health and higher life expectancy, better parenting, lower crime, more efficient administration, better environment, wider political and social participation and greater social inclusion and cohesion (see more in Part 4, Section 2.1. and Box 4.4 ⁽⁵⁴⁾). Another example for externalities is the avoidance of unemployment for better educated people, which is both a private benefit and also one for society. Moreover, it exerts an indirect economic effect on growth by reducing the burden on social insurance systems (particularly unemployment insurance) as the money saved by lower unemployment payments can be invested in more productive areas.

These effects are hardly captured explicitly within macroeconomic models of growth. Their relationship with education and skills can nevertheless be measured by econometric analyses. However, in many cases, education and training influences these issues indirectly and may also yield positive effects on

⁽⁵²⁾ To avoid linearity, most studies use a square term for schooling, resulting in an inverted-U pattern of the returns to education. Thus, additional years of schooling have a decreasing effect on productivity.

⁽⁵³⁾ As well as to individuals.

⁽⁵⁴⁾ We should note that there might also be a 'perverse causality' between these issues and economic growth, at least in the short and medium term. For example, lower health, higher crime rates and social exclusion may increase GDP because of higher incomes and more workers required in the health sector, jurisdiction, police, etc. However, since most of these activities have low labour productivity, money saved for these jobs and invested in more productive sectors will increase growth and productivity in the longer term.

economic growth. For example, some forms of crime and social exclusion may be due to an antecedent variable: poverty. One important tool in the fight against poverty is better education and training, which in turn contributes to reducing crime and improving social inclusion. Similar effects can be expected by raising the skills of public services personnel, thus rendering public management and institutions more effective and positively affecting social infrastructure, trust and citizenship, among others.

Externalities and spillovers are normally not taken into account in decisions of individuals or companies on educational investments. Their existence is, however, 'an important economic justification for the public support of education and is often assumed *a priori* by theorists and policymakers alike, although the difficulties of actually verifying their size and thus calculating true social returns are formidable' (Sianesi and van Reenen, 2002: 6).

Social rates of returns on education and training attempt to include all costs and benefits and use pre-tax earnings. They take into account private costs, public costs such as subsidies to education and income taxes. However, although ideally the social rate of return should include all costs and all benefits, they normally do not include non-material benefits or externalities which refer to diverse macroeconomic gains and benefits for the society.

Without taking into account all social benefits of education and training – on top of private ones – social rates of return tend to be lower than private rates of return. Thus, social rates of return should be regarded as a lower bound of the full return on education and training (Sianesi and van Reenen, 2002: 10) ⁽⁵⁵⁾.

2.1.2.2. *Indirect effects of human capital on growth*

Indirect effects of human capital on growth can be investigated by using regression techniques similar to those that try to identify the determinants of growth, although with different dependent variables. Some of these indirect effects are ⁽⁵⁶⁾:

- (a) in more advanced OECD countries, tertiary education investments have a direct effect on growth, while the stock of human capital at secondary level, including vocational training, appears particularly important in stimulating investments and thus has an indirect effect on growth ⁽⁵⁷⁾;
- (b) women's education and/or higher income appear to be associated with significantly lower net fertility and thus lower population growth. If GDP remains unchanged, this results in higher productivity because GDP is produced by a smaller number of workers;
- (c) educational attainment and higher income levels are associated with higher life expectancy, lower infant mortality and higher levels of primary and secondary school enrolment rates of children. This also may in turn have an indirect positive impact on economic growth if the positive link between education/training and growth holds.

2.1.2.3. *Social infrastructure and social capital*

Beyond the quantitative assessment of the direct and indirect impact on economic growth, questions arise. What are the reasons why some countries accumulate more human capital than others? What makes them more efficient than others?

One of the reasons is a well-developed social infrastructure (or superstructure; see Box 2.7), i.e. policies, norms and the legal and institutional framework which shape the environment in which people and firms act and

⁽⁵⁵⁾ A more detailed discussion on the internal rates of return, both private and social, and its links to the Mincerian rate of return to schooling is given by de la Fuente and Ciccone (2002: 44 ff.); see also Section 2.3.1.1 of this report. For a critique of the significance of rates of return for economic growth see also Wolf (2002).

⁽⁵⁶⁾ The examples that follow come from the review by Sianesi and van Reenen, 2002: 35 f.

⁽⁵⁷⁾ In contrast, in less advanced or developing countries an increase in basic skills at primary or lower secondary level has a pronounced direct effect on growth, whereas the effect of tertiary education is lower or even negative.

make their investment decisions. A good social infrastructure cares for right pricing, markets function, avoiding market failures and ensuring social equity. In such an environment, investors are able to reap the returns on their activities, be they production, skill acquisition, invention or technological transfer.

Box 2.7. Social infrastructure

In social sciences, social infrastructure or superstructure 'is the set of socio-psychological feedback loops that maintain a coherent and meaningful structure in a given society, or part thereof. It can include the culture, institutions, power structures, roles, and rituals of the society. It is that which, through conditioned behaviors (both interpersonal and situational), enforces a set of constraints and guidelines on human activity in a stable and effective fashion, such that it engenders a society's characteristic organization [...].'

Source: Available from Internet: <http://encyclopedia.thefreedictionary.com/Social%20infrastructure> [cited 1.6.2004].

A related concept is social capital which encapsulates the norms and social relations embedded in the social structure of a group of people who collaborate to achieve desired and shared goals. Social capital appears to be positively correlated with a country's economic success ⁽⁵⁸⁾ (Box 2.8).

'Social capital as a determinant of economic growth has received much attention in the last decade. It is important to understand at the outset, however, that social capital research is still at its beginnings and that it should be seen as a collection of suggestive arguments and pieces of empirical evidence, rather than as a set of conclusions that can be of direct use in the formulation of economic policy' (de la Fuente and Ciccone, 2002: 34).

According to Healy, social capital refers to aspects of social life, including the existence

of networks, norms and relationships 'that enable people to act together, create synergies, and build partnerships. Coleman (1990) showed how social capital could influence the ability to acquire human capital, for example when strong communities enhance learning at school. Social capital sets the context in which human capital can grow and have an impact. Social capital needs to be distinguished from an even wider range of social arrangements and influences that impact on learning and its outcome' (Healy, 2000: 20).

From an empirical point of view, social infrastructure and social capital cannot be measured directly. Instead, a number of proxies is being proposed, such as trust and civic norms, associational activities, networks, etc. ⁽⁵⁹⁾. There are a number of factors which influence social capital, education and training being an important one. Here again, reverse causalities may be found since several of these ingredients of social capital are strongly influenced by, or at least associated with, the economic performance of a country ⁽⁶⁰⁾.

2.1.2.4. Techniques to evaluate 'macrosocial' benefits of education and training

Green et al. (2004) propose a number of techniques and approaches to assess the macrosocial benefits of education and training. This section is based on their report.

There is no single technique to capture the whole range of macrosocial benefits, their qualitative dimension and the historical and cultural context in which they are embedded. More suited are mixed methods, involving qualitative and quantitative techniques which refer to underlying factors of influence rather than to variation between individuals.

Furthermore, the particular concept of evaluation – and in particular summative evaluation – is of limited use in this context.

⁽⁵⁸⁾ In addition, individual social capital represents the social skills and competences of a person. It enables the individual to cooperate and interact with others.

⁽⁵⁹⁾ One of the most important data sources for measuring these proxies is the World value survey (WVS), coordinated by the Institute for Social Research of the University of Michigan, under the direction of R. Inglehart (available from Internet: <http://www.worldvaluessurvey.org> [cited 9.7.2004]). The use of the WVS for analysing macrosocial benefits of education and training will be discussed in more detail in Part 4, Section 2.1.5.

⁽⁶⁰⁾ These questions will be discussed further in Part 4.

Box 2.8. Defining social capital

The term 'social capital' became popular by the studies of Coleman (1988, 1990) and Putnam (1993a; 1993b; 1995a; 1995b) ^(a).

Social capital is a difficult concept to define precisely (for the historical roots, which can be traced back to Alexis de Tocqueville, Emile Durkheim and Max Weber, see Woolcock, 1998). Most simply, it refers to the social norms, interactions and networks that facilitate collective action. 'It embraces the nature of social ties within communities, the relationship between civil society and the state and the quality of governing institutions. The development of social capital is likely to prove highly significant for improvements in economic well-being; a wider concept than simply growth in per capita GDP' (Wilson and Briscoe, 2004: Section 6.9).

Putnam (2000: 19) defined social capital as 'social networks and the norms of reciprocity and trustworthiness that arise from them'; he emphasises the role of civic engagement in fostering democracy and social cohesion. A similar concept is 'cultural capital'.

Similarly, de la Fuente and Ciccone (2002) define social capital in general as 'the norms and social relations embedded in the social structure of a group of people that enables the group or individuals participating in it to achieve desired goals'.

These definitions are somewhat distinct from the notion of 'individual social capital', which refers to an individual's (social) skills when interacting with others. These skills should be seen as a part of the individual's human capital (*ibid.*).

The OECD (2001d: 40) identifies four broad approaches to social capital:

- in the anthropological approach, humans have natural instincts for association; thus, there is a biological basis for social order and the roots of social capital lie in human nature;
- sociology describes social norms and emphasises features of social organisation such as trust, norms of reciprocity and networks of civic engagement;
- economists assume that people maximise their personal utility, decide to interact with others and draw on social capital resources to conduct various types of group activities;
- political science emphasises the role of institutions, political and social norms in shaping human behaviour. The World Bank gives social capital an important role in reducing poverty and promoting sustainable development by emphasising the role of institutions, social arrangements, trust and networks.

According to Healy (2000: 20), the OECD's notion of social capital refers to aspects of social life, including the existence of networks, norms and relationships 'that enable people to act together, create synergies, and build partnerships. Coleman (1990) showed how social capital could influence the ability to acquire human capital, for example when strong communities enhance learning at school. Social capital sets the context in which human capital can grow and have an impact. Social capital needs to be distinguished from an even wider range of social arrangements and influences that impact on learning and its outcome.'

(a) The World Bank (2002) has a website with an entire electronic library on the subject.

Green et al. (2004), with reference to Plewis and Preston (2001) distinguish between three types of activity: measuring, modelling and evaluation.

Measuring macrosocial benefits is central to any evaluation. However, measuring alone tells us little about the relationship between educational activities or systems and their benefits. For example, we cannot assume *post-hoc* that an increase in life-expectancy in a country arises due to increasing levels of education.

Some kind of modelling is required to make such statements. Basically, this may involve descriptive comparison of aggregates through scatterplots or correlations. Also used are

more advanced techniques such as multi-level modelling (MLM) or structural equation modelling (SEM), although it is debated whether they present a basis for analysing actual macrosocial aggregates rather than contextualisation of microsocial relationships (see Green et al., 2004, for more details).

Relationships over time can be modelled using time-series data on educational levels, educational distributions, macrosocial aggregates and appropriate controls. This would allow assessment of causality, for example whether changes in education cause changes in social cohesion independently of other influences. However, time series data on macrosocial and educational indicators is

difficult to obtain. In the EU, for example, measures of skill distribution (e.g. derived from IALS) or values (e.g. derived from the WVS) are only available over short time periods and causality is difficult to determine.

Another problem refers to the ‘ecological fallacy’ which occurs when using data which have been aggregated from individual surveys and represent average values of an indicator (see Box 2.9).

Box 2.9. Ecological fallacy

The ‘ecological fallacy’ is often referred to as a reason why a macrosocial analysis is inappropriate. The ecological fallacy holds when the wrong units of analysis are considered in making an interpretation of data. For example, the result that both the average level of education and the average level of tolerance are high in a given country does not indicate that there is a relationship between education and tolerance for all individuals in that country – this is the fallacy of aggregation.

It is therefore important to choose an appropriate unit of analysis. The aggregate level of trust elicited through individual trust levels may be meaningful in itself, and useful in analytical work. There are also some indicators which do not apply to the individual level such as skill distributions, income distributions, ethnic conflict, industrial disputes, social cohesion or government corruption.

Source: Green et al. (2004).

For comparative analyses, microsocial data can be used to model relationships between education and social outcomes in various countries and then to compare effect sizes. Approaches which could be used here include regression analysis or multiple comparison of groups. Also, MLM can be used to examine interactions between effects at country, regional and individual level. However, since MLM requires a sufficient number of sampling units, this technique is less efficient for comparisons between a small number of countries.

It may also be possible to monetarise educational benefits in order to calculate a social rate of return. Although most calculations of social rates of return focus only on the costs and benefits of education in terms

of expenditure and tax and income revenues, significant progress was made towards monetarising other social benefits such as intergenerational transfers, health and crime (McMahon, 2000).

The calculation of a full social rate of return, which would include the whole range of social benefits, is difficult to achieve. Many of the macrosocial effects of education, such as social cohesion and changes in attitudes and values, cannot be easily incorporated within a social rate of return. Moreover, as the macrosocial effects of education and training take effect over long periods of time, the social rate of return needs to be calculated with reference to inter-generational considerations.

Measuring and modelling, whether in terms of regression, MLM or calculation of the social rate of return, do not necessarily mean evaluation. Evaluation implies systematic analysis of the effects of a particular programme or activity which intends to inform decision-makers on the success (or not) of an intervention and eventually to modify the design of existing or future programmes. Modelling usually does not investigate the effects of a particular programme and is normally not built into programme design.

As Green et al. (2004: Section 5.1) point out, summative evaluation is of limited value in research on the macrosocial benefits of education. National education systems normally do not represent targeted programmes (see also Part 3, abstract). Although specific programmes or policies can also be evaluated in a summative sense, the evolution of education systems over longer periods of time and their contested nature make the authors sceptical about claims that they are designed to meet discrete social objectives (although there may be general aims underlying education systems). Moreover, the embeddedness of education systems within national cultures and institutional structures makes it difficult to separate out the particular effects of education.

The authors conclude that research concerning the macrosocial benefits of education is allocated somewhere between

Box 2.10. Productivity, profitability and market value – some definitions

(a) Productivity measures the ratio of the quantity of outputs (goods and services) produced to the quantity of inputs (labour, capital, raw materials, etc.) used. Productivity can be measured at the level of firms, industries and entire economies. Productivity is dependent on efficient use of resources, the introduction of new technologies and work organisation, etc. Productivity can be a physical measure (e.g. number of cars produced per employee) or a monetary measure (e.g. thousands of Euro of output per hour worked). The most-used input categories are labour (number of employees, hours worked, skills) and capital (buildings, machinery, equipment, raw material, energy, etc.).

Labour productivity is the ratio of the real value of output to the input of labour (or hours worked). This measure should be interpreted with caution because it reflects more than just the efficiency of workers. Labour productivity is influenced by many factors outside of workers' influence (e.g. capital, new technologies, management, organisation). For these reasons the productivity of all production factors (total factor productivity) is a better indicator of efficiency.

Source: Productivity Commission, available from Internet: www.pc.gov.au/work/productivity/primer.html [cited: 10.5.2004].

(b) Profitability indicates a company's earnings in relation to sales or assets. Commonly used profitability ratios are:

(i) profit margin: (earnings after taxes): sales;

(ii) return on assets (ROA): (earnings after taxes): total assets;

(iii) return on equity: (earnings after sales): equity;

(iv) price to earnings ratio: stock price: company's earnings per share.

Source: available from Internet: www.ameritrade.com/educationv2/fhtme/learning/ [cited: 10.5.2004].

(c) Market value is the price at which the property (e.g. a company) would change hands between a willing buyer and a willing seller, neither being under any compulsion to buy or to sell and both having reasonable knowledge of relevant facts. For publicly traded stock the stock market (average prices on a given day) determines the value. Other valuation methods focus on profits and assets of a company. The book value is the value of a company's assets minus its liabilities.

Source: available from Internet: www.fairmark.com/execcomp/fmv.htm [cited: 10.5.2004].

(d) Market share is the ratio of sales of a brand to the total sales of that product-type in a defined area (country, continent, etc.). Market share can also be defined as the ratio of a company's entire product line to the total sales of all related companies. Market share is usually presented as a percentage [...].

Source: available from Internet: www.lib.duke.edu/reference/subjects/business/m_share.htm [cited: 10.5.2004].

modelling and evaluation. Comparative research on the macrosocial benefits cannot remove confounding influences. 'Countries are historically produced, open systems, and it is not possible to subject education systems or nation-states to control or experiment (at least social-democratic ones). It is possible, though, to gain understanding of developmental processes and to identify similarities and differences between countries. In this way, a hybrid representing modelling and formative judgements may represent the closest comparative research of this type can get to evaluation. With this in mind, we therefore proceed to a preliminary investigation of the effects of education, training and skills on macrosocial benefits' (ibid., Section 5.1).

2.2. Methods to investigate the impact of skills on company performance

Actors in the labour market, as well as researchers, are interested in solid information on companies' investments in training – both initial and continuing – because it is expected to be one of the main reasons for company performance. Nevertheless, in contrast to macroeconomic analyses of GDP growth and productivity, the factors that influence company performance and investments – in particular of human capital – are still under-researched.

Furthermore, the impact of company investments in education, training and skills of employees on company performance is difficult to measure because of numerous intervening factors – internal and external – which affect productivity, profitability, competitiveness and other performance indicators. Moreover, business cycles, legal and other regulations and industrial relations between employers and workers – to mention a few – influence a company's production, production process, work organisation, pay scales and wages paid to employees ⁽⁶¹⁾ and training policy.

2.2.1. Training investment and company performance

From a company's perspective, investments in training bear the risk that trained people may leave the firm and another firm will accrue the benefits of previous training investments. This is a disincentive for companies to engage in training. Becker (1962) argued that this is particularly the case if firms provide 'general' training that is 'marketable'; i.e. it can also be used by other companies. In contrast, specific training only benefits the training company which consequently should bear the whole training costs.

Testing Becker's theory raises the questions of what the benefits are, who benefits and who should pay for different types of training. Investigating these issues requires not only data on wages but also, and more importantly, information about a firm's benefit and performance.

Performance measures include productivity, profitability, market share, market value (Box 2.10). Another performance measure is innovation and innovativeness using indicators such as patents, dependent among other things on R&D activities.

2.2.2. Enterprise surveys ⁽⁶²⁾

Although enterprise data are being collected systematically at national and at European

level, the aspect of training and skills is often neglected.

There are basically two approaches to collecting data at company level: cross-section and longitudinal surveys. Cross-section surveys collect company data at one or more subsequent points of time; however, companies included in subsequent surveys are not necessarily the same. In contrast, longitudinal surveys, e.g. panels, collect data on identical enterprises or firms/establishments over a time period. Both types of surveys may use a modular questionnaire with core items included in every survey and variable items (e.g. those which are not due to rapid changes) over longer time distances. Although longitudinal surveys are much more time consuming and expensive ⁽⁶³⁾, they have decisive advantages in that they take into account the heterogeneity of enterprises, cause-impact relationships and the adaptation behaviour of an enterprise.

A disadvantage of longitudinal surveys is selectivity problems caused, among others, by panel mortality. In general, the probability of survival of a newly established firm is significantly lower than of firms that have existed for some time. For this reason, the panel should be regularly complemented by new firms which should be presented above average in the longitudinal sample (oversampling).

There are a number of important aspects that require panel data. Examples are (Descy and Tessaring, 2001a):

- (a) the costs of adaptation of staff to changed requirements; these costs normally accrue in a longer time period;
- (b) the question of why an enterprise does not provide training at all or only during a certain period;
- (c) developments over time can be better analysed by repeated enquiries of identical units; this allows the distinction between changes in behaviour of a single firm and changes in the composition of an aggregate of firms;

⁽⁶¹⁾ Including voluntary and statutory labour costs and non-wage labour costs.

⁽⁶²⁾ This section is mainly based on Bellmann (2001) and Descy and Tessaring (2001a).

⁽⁶³⁾ In addition, results will be available only after a longer time period.

(d) questions concerning past periods (as used in cross-section surveys) are subject to mistaken recall gaps; this is not the case for panel data collected over a longer time period.

Other important questions to be answered when conducting enterprise surveys⁽⁶⁴⁾ concern adequate survey unit: the establishment as a local unit or the enterprise that may be a conglomeration of several establishments. Experience from enterprise surveys has shown a clear preference for the establishment/unit level since decisions on training and recruitment are increasingly devolved to lower levels in the course of reorganisation of enterprises. Furthermore, important economic indicators, such as production and turnover, working time, wages, etc., are more likely to be available at the establishment/unit level.

Finally, the addressee of a survey questionnaire within a company has to be considered carefully. Whereas in small and medium sized companies the owner or executive manager might be the right addressee, this might be different in larger companies. According to the subject field of the questionnaire, the personnel manager, accounting officer or an in-company market research executive might be appropriate persons to respond to questions on, for example, skill levels of employees, training activities, business data and development plans.

Although a large number of enterprise data are collected systematically, they do not always focus on, and may even neglect, aspects of training and skills.

2.2.3. Weaknesses in defining training⁽⁶⁵⁾

The lack of information on training investments poses considerable problems for research on the benefits of education, training and skills for a company. Data may have to be gathered from different sources, often lacking common definitions, thus hampering comparison. This

is particularly true for the definition of 'training'. For example, the Institute of Personnel and Corporate Development (IPF) at Uppsala University surveys companies listed on the Stockholm stock exchange. The human capital survey 2002 included questions on what firms defined as company training. Some companies report only training conducted outside the firm (12 %), other companies report internal and external training sessions with a defined curriculum (39 %), whereas others report anything from formal training sessions to informal training such as learning by doing and self studies (45 %).

This lack of a coherent definition of training that is used and reported consistently by companies is one of the most important concerns of research on firm-based training. If there is a true relationship between training and firm performance, vague definitions of training make the estimations less precise and less significant. Definition problems might partly explain the low or insignificant impact of training in some studies. Furthermore, this problem is not likely to be solved by surveys which predefine training because it is difficult to imagine that firms would rearrange their training data for each new survey. Some guidelines or general agreement on how to define training would be beneficial for both research and companies⁽⁶⁶⁾.

Apart from lacking a common definition and standard of training, another problem is the definition of the type of costs that should be included in training measures, i.e. opportunity costs of workers trained during their working time, costs of sideline trainers employed in the company, gross and net costs, etc. An accepted measurement of the financial investments in training will not only facilitate cross-country and cross-study comparisons but also comparisons with other types of tangible investments. If investment in human capital could be compared with, and have the same credibility as, tangible investments we

⁽⁶⁴⁾ For a detailed discussion on the requirements of company surveys concerning, for example, fieldwork, questionnaires and data processing, see Bellmann (2001).

⁽⁶⁵⁾ This section draws mainly on Hansson et al. (2004).

⁽⁶⁶⁾ An example of a general agreement on definitions and design of company surveys are the ASTD company surveys (see Part 4, Section 3.3.1).

could better understand its contribution to wealth for firms and society.

When studies use panel or longitudinal data, the definition problem is less important. Differences in training definitions between companies do not matter so much then as long as the unit of analysis (the company) does not change its own definition of training. Furthermore, if the same company is surveyed over time intervals, all time-invariant effects that can bias results are eliminated. Conclusions drawn from panel studies are therefore much stronger.

Concluding with Hansson et al. (2004), the weakness of definitions, and therefore of data used in many studies on the links between training and company performance, is understating the significance of human capital investments. Inadequacies in methods or insufficiencies of data may render true relationships insignificant and lead research on the determinants of company performance to understate the impact of human capital investments.

2.3. Investigating individual benefits of education and training

2.3.1. Private returns on investment in human capital ⁽⁶⁷⁾

Human capital theories explaining the individual demand for education were formulated in the 1960s by economists such as Gary S. Becker, Theodore W. Schultz, Edward F. Denison and Jacob Mincer. Education (and training) is seen in these theories as an investment by individuals which will be made only if future gains, or returns, are expected. Returns on education and training can be measured in monetary terms (e.g. earnings) and as rates of return, taking into account the costs of investments in education and training; these may be direct, such as tuition fees, or indirect, such as opportunity

costs (forgone earnings during the time of additional education or training). Non-monetary returns include lower unemployment period and a range of diverse non-material benefits such as better health and reduced crime.

The distinctions between direct and indirect returns on education and training, and between monetary and non-monetary returns, are also the object of many studies ⁽⁶⁸⁾.

Human capital theories assume that an individual chooses education to maximise the total stream of future earnings. In a competitive labour market, wages reflect the marginal product of workers. Better educated or trained people must be more productive than their lesser-skilled counterparts and therefore will achieve higher earnings (Blundell et al., 1999). In the absence of direct measures of an individual's productivity, higher skilled workers receive higher earnings because employers assume that they are more productive. According to this thinking, maximising lifetime earnings is (also) a result of additional education and training.

Box 2.11. Some definitions of returns and costs

'Monetary returns' (net earnings differential) are the additional returns attributable to any given level of education, before and after retirement.

'Non-monetary returns' (e.g. satisfaction, better health, lower unemployment) denote the marginal increment generated by highest level of education attained.

'Investment costs' are direct or indirect. Indirect costs are opportunity costs (forgone earnings), borne largely by the parents. Direct costs are tuition, fees, books and scholarships/subsidies, the latter borne largely by the governments in the case of public institutions and endowment funds and gifts in private schools.

2.3.1.1. Rates of return

The rate of return on investment in education and training, or in human capital, is a monetary measure which takes into account both the

⁽⁶⁷⁾ This section draws mainly on de la Fuente and Ciccone (2002) and on Wilson and Briscoe (2004).

⁽⁶⁸⁾ For a discussion and illustration of age-earnings profiles, monetary and non-monetary returns over the life cycle and investment costs see, for example, Mac Mahon (1998: 459 f.).

Box 2.12. Calculation of rates of return

There are two basic approaches to calculating rates of return: the discounted cash flow (DCF) accounting framework and the earnings function approach (see for the following Wilson and Briscoe, 2004).

(a) The DCF accounting approach compares lifetime earnings streams of individuals who participated in different education or training programmes. By discounting lifetime earnings back to the beginning of working life (i.e. after leaving the respective education and training programme), an internal rate of return on the investment is generated which makes the net present value of the future earnings stream equal zero. This rate can be compared with rates of return on other (alternative) education and training programmes or with the general rate of interest in an economy. This internal rate of return is equivalent to the internal interest rate calculated for any investment in capital, including savings. Comparing alternative rates of return allows for a valuation of investments in different human capital enhancing programmes or in physical capital.

(b) The estimation of rates of return with the earnings function is based on earnings data of samples of individual workers, together with information about their personal and other characteristics, including their investments in education and training. The rate of return expresses the value of an additional year of education or training (or the value of a particular qualification) in terms of the associated increase in earnings. This method addresses problems of endogeneity and spurious correlation directly and has become the accepted standard.

The empirical approximation of the human capital theory framework is the wage (or earnings) equation. The specification mainly used to estimate the effect of individual schooling on individual wages follows the basic approach developed by Mincer (1974) ⁽¹⁾:

$$\ln W = \alpha + \theta S + \gamma E + \mu E^2 + \varphi X + u$$

where W is the wage (or earnings), S schooling (mostly expressed by school years), E experience, X a set of other individual characteristics and u the variation in wages not observed by the variables in the right side of the equation; α , θ , γ , μ and φ are parameters. Note that 'experience' is introduced also as a quadratic term (E^2) to capture the concavity of the earnings profile ^(a).

Of particular importance is the parameter θ which measures the increase in wages correlated with an additional year of schooling. According to Mincer, θ can be interpreted as the private return on schooling or as the proportionate effect on wages caused by an increment to schooling (S).

In effect, both approaches attribute the education-related difference in earnings between individuals or groups of individuals having or having not participated in an education or training programme. This comparison is only valid if the individuals being compared (i.e. those with and without the education or training) are identical in every relevant respect. Moreover, this approach takes no account of selection biases (see Section 2.3.2.1) ^(b).

(a) See for the description of Mincer's earning equation de la Fuente and Ciccone (2002: 16). The following derivation adopts the notation used by de la Fuente and Ciccone (2002: 16).

(b) If there is no information about work experience, 'potential experience' is used as a proxy. This denotes the number of years that an individual could have gained work experience after leaving education. Most common is to calculate the difference between the actual age of an individual and the typical age when leaving education. When comparing rates of return between different levels of education, different ages of leaving education have to be assumed. Comparing the earnings profiles of these different levels indicates the 'overtaking' point; this is the age year where earnings for a particular level, by accounting for opportunity costs, surpass those of lower levels (without or with lesser earnings forgone by additional education).

stream of earnings over a lifetime and the costs of additional 'schooling' born by individuals (Wilson and Briscoe, 2004: 10). It includes the benefits of education and training in terms of post-tax lifetime earnings minus both direct costs (e.g. tuition fees) and indirect

costs of education, namely opportunity costs (i.e. earnings forgone by additional schooling). These returns on the resources invested in education and training can be computed into one single indicator, the internal rate of return ⁽⁶⁹⁾ on schooling. It is defined as the

⁽⁶⁹⁾ The internal rate of return is the discount rate which sets the net present value of the future earnings stream equal zero. An alternative measure is the net present value of an earnings stream which is the sum of the present values of the individual amounts in the earnings stream. Each future amount is discounted with a rate (e.g. the interest one would have to pay if the amount were borrowed) representing the opportunity cost of holding capital from now until the year when earnings are received. In other words, the net present value indicates how much money you would have earned if you invested the money elsewhere or how much interest you would have to pay if you borrowed money.

discount rate that makes the net present value of the accumulated lifetime net benefits (earnings stream minus costs), generated by additional schooling, equal zero (Box 2.12).

The individual can then compare the internal rate of return with a reference discount rate, e.g. of investment in physical capital, share or equity investment returns or a going market interest rate (e.g. on savings). This calculation of the rate of return follows similar principles to the calculation of the net present value within a cost-benefit analysis described above (Section 1.3.4.2). If the rate of return on schooling exceeds alternative discount or interest rates, the individual will decide to invest more in his or her human capital.

Typically, rates of return are computed for different qualifications or for the length of time spent in education or training courses. Various studies include variations in rates of return according to type of qualification, gender, age, experience and ability or investigate the effects of over-education and the direct impact of education and training on the probability of employment.

2.3.1.2. Methodological challenges

A number of problems arise when estimating the causal effect of education and training on individual earnings (for a critical discussion see Wolf, 2002; Blundell et al., 1999). The most discussed issue is whether higher earnings of higher skilled workers are caused by better education and training, or whether individuals with greater ability choose to undergo more education and training. In the latter case, 'estimates of the return to education and training will be too large, as they will be unable to separate the contribution of unobserved ability from that of education and training (so-called "ability bias").' (Blundell et al., 1999: 3).

Whereas biases caused by unobserved characteristics ⁽⁷⁰⁾, such as ability, were not taken into account in early studies on returns on education and training, more recent research puts strong emphasis on their control. If ability is positively correlated with

both schooling and wages, the effect of schooling would be overestimated.

de la Fuente and Ciccone (2002: 15 f.) distinguish between two approaches both based on the Mincerian wage equation (see above) to resolve the difficulties caused by unobserved characteristics:

- (a) the use of data for identical twins who are assumed to have very similar ability characteristics and thus avoid the problems associated with unobserved characteristics such as innate ability, motivation, race and other genetic traits as well as family background and prior schooling;
- (b) the instrumental-variable (IV) estimation which includes additional variables (instruments) that affect the years of schooling but are not correlated with omitted determinants of wages. Examples for those instruments are institutional changes affecting school leaving age or changes in tuition costs.

2.3.2. Benefits of education and training in the life course ⁽⁷¹⁾

A serious caveat of the calculation of rates of return on education and training is that future earnings associated with additional schooling either have to be forecast or calculated by comparing the additional earnings of people with different ages at a given point of time, thus assuming that these 'age-earnings profiles' will remain unchanged in the future.

This problem can be overcome only if the real development of additional earnings caused by more education/training is observed over the whole life cycle of individuals. This issue is addressed in a relatively new strand of research in education and training: life-course or biographical research. It explores the various material and non-material benefits generated by education and training over their whole life. These benefits include not only earnings but also, for example, longer life expectancy, better health and well-being, lower unemployment and

⁽⁷⁰⁾ This issue was discussed in Section 1.3.2.1 (on selectivity).

⁽⁷¹⁾ This section draws on Heise and Meyer (2004) in its main parts.

more advantageous work careers, higher pensions, etc.

The individual life course can be considered with a great variety of social theories in mind. For this research report, and following Heise and Meyer (2004), we will use the following definitions of life course and biography.

- (a) The life course of an individual is composed of – and mostly determined by – a series of decisions an individual makes at several specific points in his/her history. Many of these points are institutionally or socially determined, such as entry into compulsory education, into training or higher education, into adulthood, retirement, etc. Others are determined by individual preferences or external conditions, e.g. family or labour-market situation. Each decision influences subsequent ‘pathways’ in society and its social structure. These pathways represent typical life patterns and more or less well-defined sequences. Changes between pathways (such as status passages or social mobility) denote transition points, i.e. the exit out of one sequence and the subsequent entry into a limited number of new sequences. Although the individual life course is structured to a degree and significantly determined by social institutions, individual decisions between alternative opportunities are central to the life course and constitute the observed variety of societal patterns.
- (b) An individual’s biography is defined as the subjective interpretation of all events in an individual’s life. Biography is thus the self-perception of one’s life-history and identification with social positions. It is recognised by the individual as unity, in contrast to the clear-cut sequences of pathways determined by social institutions. As a result, typical sociocultural milieus with common interests, experiences, attitudes, knowledge, etc., are formed and have an impact on social stratification. Social-cultural milieus and their way of thinking about society are reflected in biographic reports on individual lifestyles or personal auto-

biographies which are important sources for biographic research. Decisions taken during the life course (and their results) are an important part of an individual biography. As Roberts (2002: 1) notes, ‘biographical research is an exciting, stimulating and fast-moving field which seeks to understand the changing experiences and outlooks of individuals in their daily lives, what they see as important, and how to provide interpretations of the accounts they give of their past, present and future’.

Life-course and biographical research have developed from several theoretical approaches, faculties and empirical research fields. ‘Theoretical approaches involved are for example human capital, segmentation and status allocation theories, gender and ageing theories, and theories of generational change. Faculties concerned are economy, sociology, social psychology, developmental psychology and social demography. Empirical fields are mobility research, family cycle observations, qualification and career research, etc. Despite all differences this kind of research has in common: the investigation in dynamic processes’ (Heise and Meyer, 2004).

Life-course research concentrates on institutionally or socially determined decision situations and well-defined alternatives, while biography focuses on the individual (psychological) process of perceiving, assimilating, understanding, and reconstructing reality. In this report, we concentrate mainly on life-course research.

2.3.2.1. *Life-course research*

Life-course research, using longitudinal data or event history analysis, is particularly appropriate to studying certain topics, for example:

- (a) equal opportunities cannot be analysed at one singular point in the educational career: previous circumstances and the process itself exert important influences to be considered (Meulemann, 1990);
- (b) individual aspirations as well as abilities and performances are also not stable parameters through time; they change

because of personal and societal development;

(c) transitions such as entries, exits or re-entries into, or out of, the labour market depend to a high degree on the time spent in the educational system or in unemployment ⁽⁷²⁾. Moreover, transition processes also depend on individual experiences, attitudes and, in particular, skills acquired during life. These determinants can (partly) be separated and controlled for when analysing differences between cohorts. This approach allows, for example, testing if ‘patch-work’ careers are an effect of structural social change generating unstable careers for individuals. It enables investigation of whether, and for what reasons, interrupted job careers or multiple changes of job in a life course display significant differences between cohorts/generations.

Life-course research places cohorts (Box 2.13) at the centre of analysis, making visible (and quantifiable) different life patterns between generations, including their typical pathways through education, training and employment (see below).

Box 2.13. The cohort concept

A cohort is defined as a group of people with homogeneous individual characteristics which are invariant in the course of time. Examples for those characteristics are: year of birth, year of leaving education, gender, social origin and the like. The members of a given cohort, therefore, share the same significant life events at almost the same historical date. By comparing successive generations or cohorts it becomes possible to identify social change, for example the change of educational and occupational trajectories and participation rates of different generations over time and their institutional determination.

The following research topics will be discussed below ⁽⁷³⁾:

- (a) monetary return on education and training;
- (b) education, training and labour-market participation;
- (c) education and transitions;
- (d) generational and cohort differences in educational benefits;
- (e) social differences in educational benefits.

Monetary return on education and training

One of the educational benefits most investigated in social and economic research is earnings ⁽⁷⁴⁾. Much life-course research focuses on individual monetary return on investment in education, training or skills (see also Section 2.3.1 above). These studies rely on the human capital theory, assuming that workers are paid for their productivity which depends mainly on their qualification and experience. The human capital theory implies, among other things, a perfect labour market and a rational individual who is fully informed about all options and future consequences of his/her decisions.

As early as the 1980s, Tuma (1985) developed a human capital model taking into account the life-course perspective. In her analysis of occupational careers she considers career courses based on ill-informed decisions and unbalanced labour-market conditions. She estimates a resulting risk of misallocation that can be corrected or compensated by job-specific training received by the individual and may lead again to returns from investment in education, training and skills.

Despite several attempts, few empirical investigations based on human capital theory used longitudinal life-course data to estimate long-term educational returns.

Education, training and labour-market participation

Life-course research investigates various patterns of participation by individuals in the labour market which are considered to be influenced by education and training: successful transitions to work and higher

⁽⁷²⁾ See Descy and Tessaring (2001a) for a discussion of unemployment and its duration.

⁽⁷³⁾ Results of related research will be presented in Part 4, Chapter 4.2.1.

⁽⁷⁴⁾ ‘Earnings’ or ‘wages’ are the payments a worker receives from his or her employment. ‘Income’ refers to the total revenues of an individual, including earnings and money gained otherwise, e.g. rents, dividends, interest.

inclination to participate in work (see sections below); advantageous job mobility and career prospects; avoidance of unemployment or of inadequate jobs; easier re-entry into employment and better access to continuing education and training programmes.

From a life-course perspective, success or failure at different stages in an individual's history influence subsequent successes or failures in the working career and beyond. Life-course research, therefore, investigates both immediate outcomes of education and training, such as successful job entry after education and training, as well as subsequent or indirect outcomes that are influenced by this successful transition.

Heise and Meyer (2004) note that the theories and approaches used in life-course research are as manifold the features of labour-market participation. The range runs from status attainment and segmentation theories to human capital, job search and career mobility theories⁽⁷⁵⁾. Approaches range from sophisticated models, using econometric methods, to mere descriptive research.

Education and transitions

Transitions – between education and training, between education/training and work, and within working life (e.g. between employment, unemployment, continuing training, etc.) – have a special relevance in an individual's life history. Successful transitions may represent another benefit of education and training. In modern societies, transitions are partly determined or influenced by institutional or social rules which structure the life course of individuals. Examples include the transition between compulsory school and training, first entry into a job, entry into unemployment and subsequent return to employment, and retirement.

Transitions are particularly sensitive phases in the life course because they are influenced not only by the individual, but also – and sometimes predominantly – by external factors

such as economic recessions, educational expansion, labour-market pressures, demographic change, etc. This applies in particular to early transitions because their success might be formative for subsequent careers and transitions.

A successful first entry into the labour market might influence subsequent educational benefits such as higher wages, adequate jobs, career prospects, etc. In contrast, unsuccessful job placement is likely to accumulate disadvantages in later life. Blossfeld (1989) even assumed that disadvantages in the early career, for example experienced in the transition from education to training and/or to work, cannot be compensated for in the whole occupational career later on. If this were true, efforts devoted to continuing and adult training and to lifelong learning would be jeopardised. Successful transitions in the early career are particularly important in countries with a strong emphasis on formal qualifications, highly differentiated professions and low permeability of occupational and educational pathways (Mayer, 1996).

However, it is not only the effects of early transitions that are important; successful later transitions, such as upward mobility in an organisation's hierarchy, and all form of mobility, which helps to avoid unemployment and promote career prospects, are potential benefits of education and training in the life course. Here, different forms of training such as continuing education and training, on- or off-the-job learning, workplace learning and non-formal or informal learning might yield benefits, such as upward mobility or avoidance of unemployment. Later learning thus may qualify the assumption of the determining role of early transitions mentioned above and is one of the major arguments put forward to foster lifelong learning.

Empirical research on transitions defines these as dependent variables, with education and training being independent variables. To estimate the pure effect of education and

⁽⁷⁵⁾ For an overview on several labour-market theories linked to education and training see Büchel (2000) and Descy and Tessaring (2001a: 287, Box 4.15).

training, a number of other factors of influence that are not, or only indirectly, linked have to be controlled for. Such factors include personal characteristics such as gender, age, social status, occupation or educational background of parents, and structural patterns such as region, country, general employment or economic situation, demography. However, here again, biases might occur due to the influence of unobservable characteristics such as ability.

Generational and cohort differences in educational benefits

Another topic of interest in life-course research is inter-generational mobility, which denotes social (upward) mobility between generations ⁽⁷⁶⁾. This research compares the educational and/or occupational life courses of successive generations, or cohorts (for a definition see Box 2.13). Social change becomes visible, such as the change of educational and occupational participation and trajectories, and the influence of institutional or other external factors over time.

The connection between cohort, period and age effects is illustrated as an ideal in Figure 2.3. This fictitious figure shows two cohorts born in 1945 and 1965 and their trajectories through education, training and employment until 2000. Analytically, different trajectories (e.g. participation in education or employment) can be assigned to three effects ⁽⁷⁷⁾:

- (a) the age effect refers to the age of cohort members at a certain historical date and indicates important typical transition points, for example leaving compulsory education and entering training or higher education, or leaving education and entering the labour market at a specific age;
- (b) the cohort effect shows the route of members of a given cohort through education, training and labour market and their behaviour over time, for example in terms of educational or labour-market

participation. These behavioural patterns can then be compared with those of other cohorts;

- (c) the period effect relates to exogenous framework conditions in a certain period of time (e.g. economic crises, war, high unemployment, rapid diffusion of new technologies, enlargement of the EU, etc.) which influence the life courses of all people and cohorts, more or less, according to their age, job, employment status, etc. in that period. They might affect participation in initial training of young cohorts or the transition to work and employment conditions for older people.

Age, cohort and period effects can be quantified by, for example, calculating differences between educational and labour-market participation rates for different generations. However, the distinction between these three effects is tautological in the sense that if two categories (e.g. calendar year and year of birth) are given, the third category (age) results automatically: in mathematical terms the system is over-determined.

Life-course research argues that different generations are confronted with different social and historical conditions. Even birth cohorts which are only a few years apart from each other may face different conditions in the labour market (caused by period effects) which may affect their pathways, depending on whether these conditions appear in their sensitive or less sensitive phases of life. For example, members of a cohort belonging to a weak birth group may have more opportunities to access education, training and jobs because of reduced competition with people of the same age. Thus, different cohorts may have very disparate opportunities and risks regarding their education, training and working career and therefore receive different educational benefits. Box 2.14 provides a fictitious example for the cohort approach.

Investigation of differences of educational benefits between cohorts or generations have emerged as an important research line in life-

⁽⁷⁶⁾ See for example Hradil and Immerfall (1997); Goldthorpe (1980); Featherman and Hauser (1978).

⁽⁷⁷⁾ For a detailed description of age, cohort and period effects see Chapter 3.1 and Chapter 5 in Descy and Tessaring, 2001a.

Figure 2.3. The course of two cohorts through education/training and employment

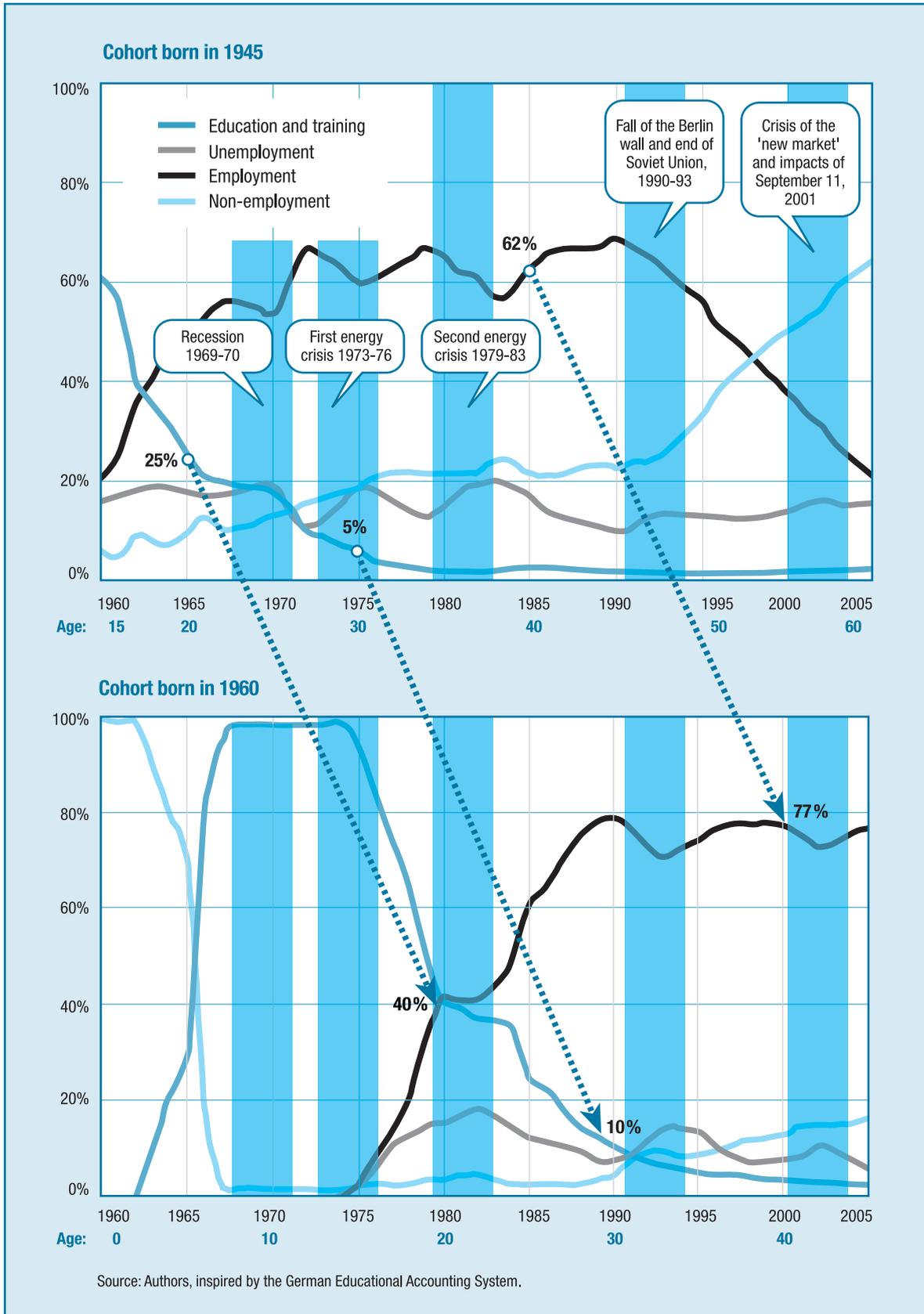


Figure 2.4. Unemployment rates in OECD countries 1965-2003



course analysis. Especially when regarding increased cross-national comparisons of educational benefits, it becomes essential to consider generational change.

Social differences in educational benefits

Research on the social determinants of educational benefits looks at the effects of ascriptive attributes such as family background, gender, ethnicity, etc. These attributes are considered as prominent factors influencing the access to, and the outcomes of, education and training.

Heise and Meyer (2004) note that although a certain equalisation of opportunities between men and women took place over the last 30 years or so, gender still matters in terms of unequal access to education, training and particular jobs. Family background is another important factor to explain differences in enjoying the benefits of education. Seen from a static perspective, children in families with a comparatively high income and social status usually have better opportunities to achieve higher levels of qualification and are

underrepresented in the group of early school leavers or in lower vocational education.

Some life-course studies assume a lack of reversibility of disadvantages experienced in early careers and argue that these inequalities are continued not only through the whole individual life course but are also reproduced in successive generations. Therefore, investigation of family background and gender differences is often combined with investigations into generational and cohort developments (Heise and Meyer, 2004).

2.3.2.2. Sources of longitudinal data

National studies of education and training benefits from a life-course perspective have been carried out mainly by statistical offices or large research institutes. The reason for this lies in the high capacity and financial resources needed for representative longitudinal investigation. Apart from France, Germany, the Netherlands, the UK, and some Nordic countries, European countries do not collect data meeting the needs of life-course analysis⁽⁷⁸⁾. Little evidence can be found for

⁽⁷⁸⁾ Even the oldest longitudinal panels (British Household Panel, German Socioeconomic Panel, Dutch Socioeconomic Panel) had their starting point in the mid-1980s. Therefore the period of measurement is limited to around 20 years covering only a part of an individual's life course. However, some panels include records on previous educational and occupational histories.

Box 2.14. Age, cohort and period effects – a fictitious example

Figure 2.3 illustrates a cohort approach for the period 1960-2005 taking the fictitious example of two cohorts born in 1945 and 1960 respectively ⁽⁸⁾. It displays the participation rates of these cohorts in education and training, employment, unemployment and non-employment. For the older cohort, the data cover almost an entire working life of 45 years, starting with the age of 15 (1960) and ending with the age of 60 in 2005. However, compulsory schooling for this cohort was mostly finished before 1960 and could therefore not be considered in the data. In contrast, the life course for the younger cohort, born in 1960, can be traced from the year of birth until the age of 45, in 2005. Therefore, both cohorts can only be compared for their age of 15-45. The age effect becomes visible when looking at, for example, the transition from education/training into working life which, in most cases, takes place at the age of 15-25. However, there are significant differences between the cohorts. At the age of 20, for example, 55 % of the older cohort was already in employment, and 25 % still in education/training. Of the younger cohort at the same age, 42 % were employed but 40 % still in education and training. Overall, participation rates both in employment and in education/training are higher for the younger cohort, with accordingly lower participation in non-employment (e.g. housewives and housemen). This significant rise of participation in education and training, and subsequently in employment (which is closely associated with higher levels of education and training) is an apparent element of social change which leads younger cohorts to undertake education and training more frequently and to stay longer in the educational system.

The cohort effect becomes visible, in our fictitious example, when looking at the course of a particular cohort through working life. Undertaking compulsory education, participation in further education and training – including higher education – and transitions to work and afterwards into retirement

become visible, cohort by cohort. Here, however, there are considerable differences between cohorts, for example the longer stay of the younger cohort in education and training, and its lower inclination to be non-employed. Thus, for example, at the age of 30, only 5 % of the older cohort participated in (continuing) education and training, the figure for the younger cohort being 10 %. Equally, at the age of 40, 77 % of the younger cohort were employed (in 2000) compared to 62 % of the older cohort (in 1985).

The period effect affects all cohorts to a more or less pronounced degree. In our example, we will take unemployment as a manifestation of the period effect (Figure 2.4); periods with increased unemployment are shadowed. The impact of the first energy crisis in 1973-76 affected predominantly the older cohort, because most of those born in 1960 were still in education. However, the second energy crisis in the early 1980s led to increasing unemployment and decreasing employment for both cohorts. Equally, the employment boom and subsequently rising unemployment after the fall of the Berlin wall and the end of the Soviet Union in the early 1990s affected both cohorts, although again to different degrees. Finally, both cohorts were affected by the crisis of the 'new market' and the repercussions of 11 September, 2001.

It becomes obvious, when trying to analyse these different life-course patterns that all effects have to be taken into account and to be controlled for. Thus, for example, the comparison of employment rates at a given date makes little sense without considering the age and status of a cohort. Equally, comparing unemployment rates for people of different ages has to take into account the overall economic and social situation as well as previous education and training pathways.

⁽⁸⁾ A cohort analysis should, however, cover all cohorts in between and beyond the two cohorts mentioned here.

southern or eastern European countries. Therefore, comparability of results at European level is rather limited. Some examples of these data sets are presented in Box 2.15 ⁽⁷⁹⁾. Given the heterogeneity of approaches and concepts, this section will present selected studies.

The scarcity of European longitudinal data limits cross-national life-course research in

educational benefits to a few studies, which are based on either European panel data, corrected or harmonised national data, or data collected or reproduced in the framework of single cross-national projects (see Box 2.15).

The only available European longitudinal data set is the European Community Household Panel (ECHP; see Box 2.16) ⁽⁸⁰⁾.

⁽⁷⁹⁾ Additional information on longitudinal studies and data sets was presented in Descy and Tessaring, 2001a: 335 ff.

⁽⁸⁰⁾ Studies on educational benefits based on the ECHP are, for example, Deding and Dall Schmidt (2002); Pedersen, and Dall Schmidt (2002); Brunello (2001b).

Box 2.15. National data sets for longitudinal analyses of the impacts of education and training

Among the most utilised data sets for secondary analysis of ET benefits are the British, Dutch and German representative Panels (BHPS, SEP, GSOEP), the British Cohort Studies (NCDS, BCS70, YSC) and the German Life History Study (GLHS). The Dutch Cohort Studies (Brabant, GLOBE) and Swedish and Norwegian panel and cohort data sets (MS, MLS, NORIS) are also often used for country specific analyses.

In total, there are around 60 different usable studies dealing in the broadest sense with benefits of education and training from a life-course perspective. Compared to cross-section analysis of individual data this might be a small number but even so the variety of conceptions, specific investigation interests, definitions of explanatory and depending variables, sample sizes and above all results makes it very hard to provide a comprehensive overview.

NB: Abbreviation are explained at the beginning of the book.
Source: Heise and Meyer (2004).

The potential of the ECHP for comparative research in educational benefits is not fully exploited. Similar to most national panels it contains only limited information on education and training, mainly on years and levels of education. Nevertheless, the ECHP allows analysis of potential benefits such as employment and CVT prospects, mobility patterns and income growth. However, the ECHP was stopped in 2001 and, therefore, available data cover the period 1994 to 2001. This corresponds to a short period of time to investigate life courses.

Other cross-country comparative studies use harmonised or adjusted national data and compare variations in educational benefits in the life course⁽⁸¹⁾. They take into account differences in the occupational and educational systems of two or more European Member States. However, the significance of

most of these findings is hampered by incomparable or rather highly aggregated data which do not allow in-depth analyses, for example based on individualised data sets.

This gap is partly closed by increasing efforts to develop cross-national and international research projects which produce additional data sets or integrate existing national data on education and training and other social issues⁽⁸²⁾. Although few European social research projects apply a particular life-course perspective, some recent projects relate education and training to different aspects of current and future life success. These studies give a more qualitative insight into the mechanisms of education and training benefits than the mere analysis of levels and years of education and quantitative outcomes in terms of higher income or job stability.

2.3.2.3. Individual biography

The ‘objective’ stance adopted by life-course research to investigate the impact of education, training or skills has been criticised because it widely ignores individual perceptions and, therefore, the subjective significance of learning and occupational biographies. The valuation and subjective interpretation of an individual’s educational biography are likely to have important impacts on this individual’s working and life career. (Heise and Meyer, 2004).

‘The biographical perspective is subject oriented and concentrates on the differences and specialities in individual learning and occupational pathways. Therefore studies of subjectively perceived benefits in individual biographies are becoming increasingly interesting and help supplement rather objective life-course notions. Furthermore, [...] biographical research which is concerned with individual education and training benefits often includes “soft” or non-material beneficial aspects and thereby complements life-course research on “hard” material education and

⁽⁸¹⁾ Examples are Hillmert (2001); Sackmann (2001); Brunello et al. (2001a); Kaiser and Siedler (2000); Brunello and Comi (2000); Müller and Shavit (1998); Brauns et al. (1997).

⁽⁸²⁾ E.g. Kieselbach, 2000; Sofer, 2000; Hannan and Werquin, 2001; Blossfeld, 2000; McIntosh and Steedman, 1999; Furlong and Hammer, 2000.

Box 2.16. The European Community Household Panel (ECHP)

The European Community Household panel (ECHP) was an annual EU-specific survey that provided comparable microdata concerning the living conditions of private households (sample size: more than 60 000) and individual persons (more than 130 000) throughout the EU Member States. The same set of individuals and households is followed over time. Covering income, employment status, housing, healthcare, education, poverty and social exclusion, and other social issues, these EU-wide social data are collected in a harmonious structure that enables them to be efficiently integrated into comparison studies. The ECHP was created and maintained by the Statistical Office of the European Communities (Eurostat), in close consultation with the Member States.

After a total duration of eight years (1994-2001) Eurostat, together with the Member States, decided to stop the ECHP and to replace it with a new instrument, EU-SICL (Statistics on income and living conditions) ^(a).

Direct access to the ECHP original data is restricted. Due to the strong demand for access by researchers and other users, however, Eurostat has developed an 'anonymised' version of the database that complies with data confidentiality requirements: the European Community household panel users' database or ECHP UDB. The ECHP

UDB is a user-friendly database and accessible to those who sign a research contract with Eurostat, stipulating strict conditions of data access and use.

The ECHP UDB microdata are structured in a series of eight survey 'waves', from Wave 1 (1994) to Wave 8 (2001). Countries covered in Wave 1 were the EU12 Member States at that time, with subsequent waves including data for new countries joining the EU (Austria – Wave 2: Finland – Wave 3). From 1997, comparable data extracted from the Swedish living conditions survey are also available.

A description of the ECHP UDB (containing a data dictionary and the codes and labels of the individual variables, and the 'anonymisation criteria' that have been applied) is given in the document *ECHP UDB – A description of variables*, available from Internet: <http://forum.europa.eu.int/Public/irc/dsis/echpanel/home> [cited 12.7.2004].

Document *PAN170: Research contracts using ECHP data*, also available from the address above, provides research projects/publications based on the ECHP data. It is constantly updated.

^(a) Original plans foresaw the launch of the full EU-SICL survey in 2003. The EU-SICL includes both cross-sectional and longitudinal dimensions.

Source: Eurostat, 2001b and 2003.

training benefits.' (Heise and Meyer, 2004: Section 4.1.6).

Biographical research is a wide field and includes approaches in sociology, anthropology, psychology, etc. It is concerned with personal narratives or 'stories' of people. Methods used are mainly qualitative inquiry; personal biographies can be regarded as individual case studies (see for more details Patton, 2002: 132). They include observation, interviews, informal discussions, review of people's projects and documents, and records, including evaluation records and tests.

At the heart of narrative analysis is the interpretation of narratives (Denzin, 1997) and the investigation of basic types of problematic moments in individuals' lives such as critical incidents, transitions and crises.

Another issue of biographical research is to highlight 'the collective experiences of people's lives by locating biographies within a

socioeconomic and political context [...]' (Merrill, 2002: 1).

The wide field of biographical research cannot be reviewed in more detail here. Readers interested in this subject are referred to the numerous databases and documentation which exist for various disciplines, for example <http://manta.colostate.edu/howto/> or the Center for Biographical Research in Hawaii, which also publishes a journal (<http://www.hawaii.edu/Biograph>). An important source for European discussion in this field is the European Society for Research on the Education of Adults (ESREA) which coordinates a life history and biography network and has organised yearly conferences on biographical research in Europe (more in: <http://www.esrea.org>).

2.4. Discussion

The benefits of education, training and skills, or of human capital, extend beyond economic or monetary aspects, though these are important issues to justify public and private expenditure in education and training. Moreover, the presence of external effects which are to the benefit of other people, companies or society in general are – if positive – an important additional justification of public educational investments.

2.4.1. Education, training and economic growth

New theories of endogenous growth offer an interesting approach to investigating the contribution of education, training and skills to economic growth. Human and social capital and R&D are seen as key elements in creating, using and disseminating new knowledge and ideas. They make education and training the main determinants of growth in the long-term. Thus, to foster growth and productivity, education, training and learning, provided or facilitated by the state or companies, should be at the heart of policies. However ‘whatever education does for growth, it is unlikely to do quickly’ (Temple, 2000: 41).

Another important issue discussed in research on endogenous growth is whether the accumulation of human capital (i.e. investments and participation in education and training), or human capital stock (i.e. the skill levels of workers) is the most important determinant of growth and productivity. This distinction reflects two major strands of endogenous growth approach. In the framework of human capital accumulation, a subsidy to education or an educational programme which raises the level of human capital will have a limited once-and-for-all effect on GDP growth. In the second approach, the growth rate of GDP will be increased forever, via dissemination of knowledge and technology. The first approach would imply policy actions to upgrade human capital in an economy, whereas the second would also require policies to foster R&D and

technology. From the current state of research, one cannot conclude what policy should be preferred and when.

However, there are many caveats associated with growth models. Problems refer to measuring human capital which is most often restricted to formal education and training and qualifications. Other problems – particularly in cross-country comparisons – are the poor data quality and the question whether it is education and training which fosters growth, or whether it is growth and economic performance of a country which leads to more education and training. Furthermore, when interpreting results of growth models, there is evidence of a non-linear relationship between education and training on the one hand and growth and productivity on the other. This means that the impacts of education and training will level out after some years. Finally, endogenous growth models assume that R&D activities, and thus long-term growth, depend on the availability of research workers, an issue which is not always present when we consider serious skills shortages in this sector. Furthermore, individuals (and firms) are likely to underinvest in education and training because they do not value or reap the wider benefits. Therefore, improving transparency on the benefits for all agents in an economy is likely to foster education and training investments and thus economic growth.

2.4.2. Social benefits and social infrastructure

External effects which spill over from skilled workers or R&D activities to other people, companies or society can be direct or indirect outcomes of education and training. In many cases, such as criminality reduction, social cohesion, trust, etc. the effects of education and training are of indirect nature, via, for example, reducing poverty or increasing the efficiency of public services. Empirical research on the wider social benefits of education and training at macro or micro level has developed considerably in recent years, supported by the availability of large world-wide surveys (such as the World value surveys) which include a number of important

indicators for social benefits, such as trust, tolerance and crime behaviour.

In this context, the notion of social infrastructure becomes an important issue because a 'good' social infrastructure (and, closely related, social capital) exerts an important influence on economic growth and efficiency, a context in which human capital can grow and have an impact (Coleman, 1990, cited by Healy, 2000).

There are several ways of measuring social infrastructure and social capital. They use quantitative and qualitative techniques to measure, model and evaluate these concepts and the links between education, training and social outcomes, and to investigate the full social returns on education and training.

2.4.3. Impact on company performance

Economic growth denotes the growth in production of all goods and services in a given country and period of time. These goods and services are mainly produced by companies. However, in contrast to the large body of research on macroeconomic determinants of growth, the factors that influence companies' performance and production growth – and in particular the role of education, training and skills – are significantly under-researched. Existing studies struggle with missing or inappropriate data, various – and in many cases not comparable – definitions of training, costs and performance measures and restricted information on the development of companies over time.

The main methods of carrying out company surveys – cross-section and longitudinal surveys – have both advantages and disadvantages. Cross-section surveys are less expensive and can include more than longitudinal surveys or enterprise panels. The latter, however, are able to take account of heterogeneity of companies and their development over time. Cause-effect relationships, for example between training investments and outcomes, and changing behaviour of firms in terms of adaptation to new challenges, can only be explored by longitudinal data.

The weakness of data and methods used in many studies on the links between training and

company performance understate the role of education and particularly training in increasing productivity and competitiveness and in generating profits. This is confirmed by findings in several studies presented in Part 4, Chapter 3 of this report. They include research results on why companies are increasingly willing to invest in general education and training, contradicting traditional theory that companies only finance specific training which is not 'marketable' and cannot be used by other companies.

2.4.4. Individual benefits and life-course research

Investigating individual benefits of education and training was, until recently, the domain of microeconomic and econometric research studies relying on human capital theories. In most cases, benefits of education and training are material, in particular monetary returns (earnings), rates of return and some forms of non-monetary returns such as unemployment probability, occupational career, etc.

A recent strand in education and training research at micro level is life-course and biographical research. These approaches are expected to overcome some of the major problems of human capital research outlined above. Life-course research focuses on a series of decisions individuals have to make at specific transition points in their life history which influence subsequent pathways. Biographical research focuses on the subjective perception of events in life which have an important impact on social stratification processes.

Many questions on education and training can only be answered by longitudinal data used in life-course research, or by interpreting biographical perceptions of individuals. These include issues of equal opportunities, transition behaviour, abilities and performance which are not stable over time but change in relation to external factors such as the labour-market situation, provision and delivery of education and training, the economy, and changes in social and political environments.

Most research studies on the diverse benefits of education and training in a life-course perspective focus on participation in

education and training and the labour market (including unemployment spells) and on occupational career and status changes. However, many factors intervene during the life course of an individual and thus render the attribution of single effects of education and training difficult after some years.

Transitions into further education/training and work are the heart of life-course research. Transitions are sensitive phases in a person's life and are also influenced by factors beyond the control of an individual. Moreover, unsuccessful transitions – in particular in the early career – may accumulate disadvantages in the further life course. This is particularly true in countries with strong emphasis on formal qualifications, differentiation of education/training pathways and low permeability and reversibility of previous decisions. If this is the case, the success of public and private investments in continuing education and training and in lifelong learning would be jeopardised.

A related strand of research is the cohort approach, i.e. observing the pathways of

people with equal starting positions (e.g. year of birth or year of leaving compulsory school) over a longer period of time. By comparing the participation of different cohorts in education, training, employment and unemployment and the changes over time, social change becomes visible and quantifiable. Research on intergenerational mobility and on the factors that influence participation behaviour are important subjects for cohort approaches.

Another issue related to life-course research is social differences which influence access to, and outcomes of, education and training. Gender, social origin and ethnicity are among the major characteristics which are taken into account in this research. The importance of life-course research in explaining social differences becomes obvious by the fact that inequalities, for example between men and women, social classes and ethnic groups change over time. Moreover, research should also attempt to answer the question whether inequalities continue also beyond the life course of individuals, by the reproduction of inequalities in successive generations.

Evaluation of education and training in a changing European context

Abstract

In this part of the report, we will discuss how evaluation can support implementation and assessment of long-term system reform and more targeted and focused programmes, interventions or projects.

‘Reforming’ a (V)ET system means taking actions that will change major parts or elements (!) in a such way that it will influence how institutions operate and interact. Reforms transform established institutional structures, plus the habits, routines and roles of the various actors. Successful reform, therefore, requires acceptance and understanding from the different actors and institutions involved. It is a process that has long-term implications and impacts. Additionally, VET reforms cannot/should not be undertaken in isolation from the other related systems (e.g. production). Examples of VET reform include implementing a new qualification structure, new equivalences and access requirements between pathways, work placement during training, a new curriculum or VET programme, etc.

Introducing programmes means developing specific actions or interventions of a limited duration, aimed at solving specific problems, achieving specific objectives or targeting specific groups. Programmes are not designed to interact with other interventions or institutions, or to influence them. They are not meant to change the functioning of institutions (although we will argue that, in reality, they do so). Programmes can be implemented at local, regional, national or international/EU level. Typical examples are training programmes and reintegration measures developed in the framework of active labour-market policies which aim to combat unemployment and other forms of exclusion.

Viertel et al. (2004) admit that the distinction between a programme or system reform may be somewhat arbitrary. A reform may be realised through several programmes, and particular programmes can be made up of several projects or specific interventions. This is why we will use the term ‘policy’ to qualify both reforms and programmes.

The traditional way of evaluating policies is to focus strictly on the policy as a separate entity which is self-contained and independent from the environment, the context and existing institutions. International literature suggests, however, a trend away from evaluations focusing strictly on single projects or programmes. A significant shift is observed towards target-oriented approaches in labour-market policy and to a systemic approach in education and training. This is caused by the restricted contribution of single projects or programmes and the limited impact they can have even if highly sustainable (Viertel et al., 2004). Grubb and Ryan (1999: 109-119) also argue for a systems perspective that would encourage thinking not about individual projects but about widely available programmes that are linked to one another and institutionalised.

(!) Later we will refer to building blocks.

'Most evaluation methods are best suited to assessing the effectiveness of individual projects or programmes' (Grubb and Ryan, 1999: 3). This reinforces the tendency to design policies that are limited in scope and target groups and that cannot do much for a country's growth and development. A systems perspective that emphasises a more coherent approach to VET policies, where the development of programmes and projects is seen in a context where they relate one to another in addition to their individual effectiveness, is needed.

Policy-makers are expanding their views regarding the nature of useful evaluations and increasingly want information regarding systemic performance. They want 'advice related to future policy directions rather than just support with the design of a certain project' (Viertel et al., 2004).

Therefore, while reviewing in the coming chapters the main approaches to evaluating education and training, we will advocate a systemic approach to evaluation and also, more generally, to policy formulation in education and training and VET.

This part of the report comprises four chapters.

Chapter 1 sketches the transformation in economies, labour market and societies that constitute the main drivers for VET change in European countries. It reviews indicators and trends such as demographic changes, population education level, skill renewal through adult learning and continuing vocational training (CVT), and employment and unemployment patterns. It also discusses briefly the labour market consequences of the progressive move towards knowledge-based societies.

Chapter 2 deals with principles and models developed to evaluate education and training and VET systems. The chapter starts by presenting a brief panorama of the main reforms that are currently being implemented throughout Europe. It then discusses in depth education and training system evaluation. Some principles and models that can be applied to education and training and VET evaluation are extracted from concrete practices in different countries or international organisations.

Chapter 3 presents system evaluation and policy learning in action. It discusses concrete examples of evaluation in the light of the principles and models presented in Chapter 2. We have selected examples of evaluation in international contexts, such as the OECD policy review or benchmarking in the EU, and at national level in Denmark and the Netherlands. The role of teachers, trainers and education and training professionals in supporting change is discussed using the example of Croatian VET reform. The chapter finishes by drawing some lessons from selected evaluations carried out in the framework of EU Phare VET programmes and the Leonardo da Vinci programme.

Chapter 4 first reviews active labour-market policies (ALMP). For about 20 years now, they have been one important way of addressing unemployment issues. Various policies have been designed, of which training is a prominent feature. In this respect, we discuss whether training is a successful and efficient policy to combat unemployment, compared to other types of ALMP. The chapter then discusses policies that have been designed at EU level to fight unemployment, in particular the European Employment Strategy (EES), its evaluation and some of its achievements. We then make recommendations for the design of ALM training as well as for its evaluation. We underline the need for evaluation research on ALMP to develop an insight into 'why' training works (or not). Finally we advocate a new approach to ALMP and training programme evaluation: a system approach that would evaluate the effectiveness of a programme in the context of its relations, interactions and complementarity with other programmes, existing institutions and main stakeholders in VET, employment and production.

Table of contents

1. Changing contexts and drivers for change in education and training	108
1.1. Introduction	108
1.2. Impact of demographic change	108
1.3. Population educational level	110
1.4. Skill renewal through adult learning and CVT	112
1.5. Educational profile and situation in the labour market	116
1.6. Towards the knowledge economy: transformation of labour markets	117
1.7. Employment and unemployment patterns in EU15 and accession countries	119
1.8. Conclusions	122
2. Principles, models and practices for evaluating education and training	122
2.1. Introduction	122
2.2. Renewing education and training policies	123
2.2.1. Coping with new contexts and challenges	123
2.2.2. Types of reform	125
2.2.3. Conclusions	127
2.3. Systemic analysis: a key to policy evaluation in VET	127
2.3.1. Internal consistency	128
2.3.2. External consistency	128
2.4. Evaluation, at the service of policy-making in education and training	130
2.4.1. Understanding and implementing change: institutional and organisational reforms	131
2.4.2. Using building blocks to manage dialogue	133
2.4.3. Looking for explanations	134
2.5. Assessing education and training systems and reform strategies	135
2.5.1. User questionnaires and surveys	136
2.5.2. Benchmarking	137
2.5.3. Macro level	139
2.5.4. Evaluation timing: short-term evaluation for long-term change	140
2.5.5. Interdisciplinarity: grasping the complexity	140
2.6. Conclusions	140
3. Policy evaluation and policy learning in action	142
3.1. Introduction	142
3.2. Evaluation in an international context (the OECD policy reviews)	142
3.3. Learning by comparison in the EU	146
3.4. Evaluation of reforms of national VET systems	148
3.4.1. The Danish case: a 'reflective gathering of experience'	148
3.4.2. Summative evaluation and VET system reform: the Dutch case	153
3.4.3. Learning from the Dutch and the Danish cases	154
3.5. The forgotten army: teachers, trainers and VET professionals	156
3.6. VET community programmes: case studies	158
3.6.1. Evaluation of Phare VET programmes in the Czech Republic, Slovakia and Bulgaria	158
3.6.2. The Leonardo da Vinci programme: selected case studies	161
3.6.3. Conclusions and ways forward for education and training Community programmes	164
3.7. Conclusions: evaluating and learning	167

4.	Combating labour market exclusion: does training work?	168
4.1.	Introduction	168
4.2.	Active labour-market policies	169
4.2.1.	Improving effectiveness and efficiency of ALMPs	172
4.2.2.	ALMP evaluation results: what works?	174
4.3.	Fighting unemployment at EU level	181
4.3.1.	The European Employment Strategy (EES)	181
4.3.2.	Evaluations carried out for the European Social Fund	183
4.3.3.	Conclusions	184
4.4.	Recommendations for ALM training evaluation	184
4.4.1.	Broadening evaluation scope	185
4.4.2.	Opening the black box: the question of why?	186
4.4.3.	Towards a systemic approach for evaluating ALMPs	187
4.5.	Conclusion: continuous improvement of employment measures and policies	189

List of tables, figures and boxes

Tables

3.1.	Recommendations made to countries as a results of the OECD thematic reviews on adult learning	125
3.2.	Stages of a thematic policy review	144
3.3.	Levels and elements of the Danish legislative framework for the Reform 2000	149
3.4.	Evaluation of the Dutch reform, 'WEB', summary table	155
3.5.	Some figures for LdV I mobility projects in Germany (1995-99)	164
3.6.	Composition of active labour-market expenditure by type of action, EU15, 1999 (%)	171
3.7.	Lessons from the evaluation literature on the effectiveness of ALMPs, composite studies	178
3.8.	Selected results from national reviews of Employment policies – EES, impact evaluation (focus Pillar I: employability, in particular training measures)	180
3.9.	Summary of main finding of ESF evaluations, Objectives 1, 3 and 4	183

Figures

3.1.	Average population and baseline scenario, EU15	109
3.2.	Education level of the population, 25-64 years old, EU15 and accession countries, 2002 (%)	111
3.3.	Education level of the EU15 population by age groups, 2002 (%)	113
3.4.	Education level of the accession countries population by age groups, 2002 (%)	113
3.5.	Participation in continuing education and training of the population aged 25-64 years by level of educational attainment, 2002 (%)	114
3.6.	Training enterprises, EU15 and accession countries, 1993 and 1999 (% of all enterprises)	115
3.7.	Participation in CVT courses, EU15 and accession countries, 1993 and 1999 (% of employees, all enterprises)	115
3.8.	Long-term unemployment rates, 25-64 years old, by educational level, 2002 (% , countries sorted by ISCED 0-2)	117
3.9.	Development of unemployment rates, EU15, AC, US, 1976-2003 (%)	119
3.10.	Work status of persons aged 15-64, EU15 and accession countries (AC) (2002)	120
3.11.	Development of unemployment rates, comparison of young people and adults, EU15, 1995-2002 (%)	121
3.12.	Systemic levels within VET	132
3.13.	Spending on active and passive labour-market policies, EU15, 1985, 2000 (% of GDP)	169
3.14.	Public expenditures on ALMPs, EU15, 1999 (% of GDP)	171

Boxes

3.1.	The 2003 Eurobarometer on lifelong learning	116
3.2.	Lack of internal and external consistency while implementing reforms of VET in selected PHARE countries	129
3.3.	The 'building blocks' approach	134
3.4.	Applying benchmarking in Denmark	146
3.5.	The open method of coordination	147
3.6.	EU benchmarks for education and training	147
3.7.	Developing policy-decision methods and accountability in the European Commission	165
3.8.	Data on labour-market policies	172
3.9.	Selected evaluations of ALMPs, with a focus on training, in some European countries	176
3.10.	EES Peer review programme on ALMPs	182

1. Changing contexts and drivers for change in education and training

1.1. Introduction

Before discussing new education and training policies being implemented across Europe, and their evaluation (Chapter 2), this first chapter will look at the contexts and challenges that require new policies. We will discuss the drivers for change for education and VET. These can come either from demographic changes and changes in the skill level of the population or from new demands imposed by the knowledge economy on the workforce, enterprises and the labour market in general. This chapter concludes with a brief review of employment and unemployment patterns. The comparison of the situation in the EU15 and in the 10 new Member States⁽²⁾ is given a special attention.

European labour markets are faced with changes, both in supply and demand. The supply side corresponds to the population available for work and its characteristics in terms of age, gender, qualifications and level of education, etc. The demand side corresponds to the number of jobs available and their profiles, for example by sectors, occupations, skills and competences required.

Mismatches between supply and demand in the labour market can be quantitative (unemployment, shortages, overtime, etc.) or qualitative (overqualification, under-utilisation of skills, precarious employment conditions, etc.) (Descy and Tessaring, 2001a: 227-314; Tessaring, 1998: 33-63). Policy interventions try to address these mismatches. They can take the form of either short-term targeted programmes, aimed at quick reintegration of people into the labour market, for example, or more profound and long-term system reform.

These interventions also aim to address various social issues. Evaluation is a tool both to accompany policy implementation and to assess whether the problems in question have been addressed.

1.2. Impact of demographic change

Changes in the structure of the population result not only from the fall in birth rates and the extension of life expectancy (ageing), but also from emigration and immigration. These are important variables in planning education and training, lifelong learning and labour-market policies. The following indicators illustrate the prospects for the EU and try to evaluate the impact of EU enlargement on demographic scenarios.

Figure 3.1 clearly shows the ageing of the EU15 population by 2040. This will be caused by the declining number of young people while the number of people aged 65 and older increases sharply (it is estimated that the proportion of older people among the total population will increase by 12 percentage points). The consequence will be a reduction in the working age population (aged 25-64)⁽³⁾.

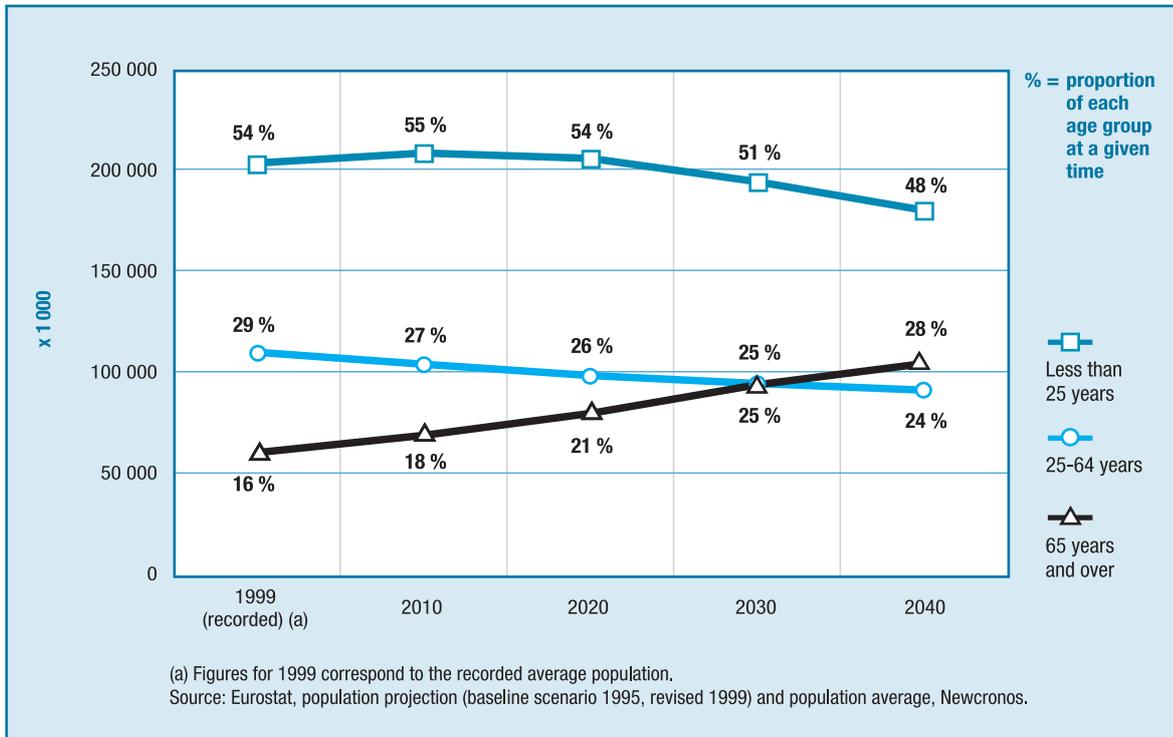
Within the working age population (aged 25-64) the proportion of 'older' (potential) workers (45-64) compared to 'younger' (potential) workers (25-44) will change. Taking as a basis the EU15 average, the current ratio of 0.79 old worker for one young one is estimated to become 1.09:1 in 2040 (Table A4 of Annex 3 for data by country).

The potential for replacement of working age populations will vary according to the proportion of young people in the population. Generally speaking, the proportion of people under 20 in the population is higher in accession countries than in EU15 countries (Figure A1 of Annex 3).

⁽²⁾ Where the data presented and the comparisons made refer to the preaccession period, we will use the term 'accession countries'.

⁽³⁾ The working age population will decline by 2040 in 12 of the EU15 Member States. The strongest reduction is expected in Italy where the number of 25-64 years old will decrease by 25 % between 2000 and 2040. A significant increase in the working age population is to be expected only for Ireland and Luxembourg (see Table A2 of Annex 3).

Figure 3.1. Average population and baseline scenario, EU15



Nevertheless, in 2001, after calculating demographic scenarios on the consequences of accession to the EU of the 12 candidate countries (i.e. also including Bulgaria and Romania at that time), Eurostat concludes: 'The accession of the twelve countries will considerably increase the total population of the Union. However, far from reversing its expected population decline, it will on the contrary hasten it. [...] On the basis of Eurostat assumptions for the Union and the expectations of the UN for the twelve countries considered, this would mean that the population decline of this group of 27 countries would start around 2015, i.e. some eight years earlier.' (Eurostat, 2001a).

In 2003, the European Commission (EC, 2003b) confirmed that the younger age structure of accession countries would first have a rejuvenating effect for the enlarged Union. However, by 2020, the proportion of older people in the new Member States will become similar to those of the EU15. In the

medium- and long-term, the accession of the new Member States will tend to reinforce the decline of the European population.

However, although the internal factors of population growth within the EU tend to become negative, international migration might rapidly become a major source of demographic increase. During the last five years, 70 % of EU15 population growth can be attributed to migration (EC, 2002a: 11). In 1999, 13 million people (3.4 % of the EU15 population) were third-country nationals; this represents an increase of 50 % over 1985 (ibid.: 18).

Building scenarios to answer the question of whether migration can counteract the ageing in the EU (EC, 2002a: 25), it appears that population size only grows significantly in 2050 if fertility rates exceed 1.8 (compared to 1.4 today) and if annual net migration exceeds 1.2 million (4). Therefore, migration probably cannot stop or reverse the population ageing process but it could contribute to filling certain

(4) This scenario takes current population in the EU15 as a basis.

specific gaps or shortages in the labour market.

The relative ageing of the population of the EU has impacts on education and training at different levels:

- (a) the reduction in young people will have impacts both on enrolment in education and training systems and on workforce skill renewal to be expected from new generations of workers ⁽⁵⁾;
- (b) continuing training of the labour force will have to be intensified to keep workers' skills up-to-date, to increase motivation and retention rates, and to ensure the quickest possible reintegration of the unemployed in the labour market;
- (c) pressures on social security and pensions funds will intensify and will increase the need to maintain people longer in the labour force ⁽⁶⁾. Measures to integrate inactive people in the labour force, in particular women, or to keep people longer in employment (as opposed to the early retirement trend) need to include a training component;
- (d) constraints created by the ageing population on public budgets may also affect education and training, since a considerable part is financed either entirely or partially by public funds;
- (e) increased life expectancy creates a progressively larger group of retired people who are active later in life. The demand for health care, social services, housing and services such as leisure, tourism, etc., will increase. VET systems should, therefore, be prepared to provide more training in the relevant occupations (Tessaring, 1998). In addition, older people will constitute a new market for lifelong learning.

A shift has to take place from traditional (V)ET provision, focused on initial education and

training, to developing continuing vocational training (CVT) and learning opportunities throughout life. This can be achieved via policies aimed at fostering private and public initiatives and provision. Active labour-market policies with a training component and adult education and training, in particular for people not in employment, should be enhanced. Given that education and training budgets are not likely to see a major net increase, funds will have to be shifted and greater efficiency ought to be achieved. In addition, the adult education and training market should adapt to more diversified needs. Finally, training in occupations related to services for older people needs to be developed.

1.3. Population educational level

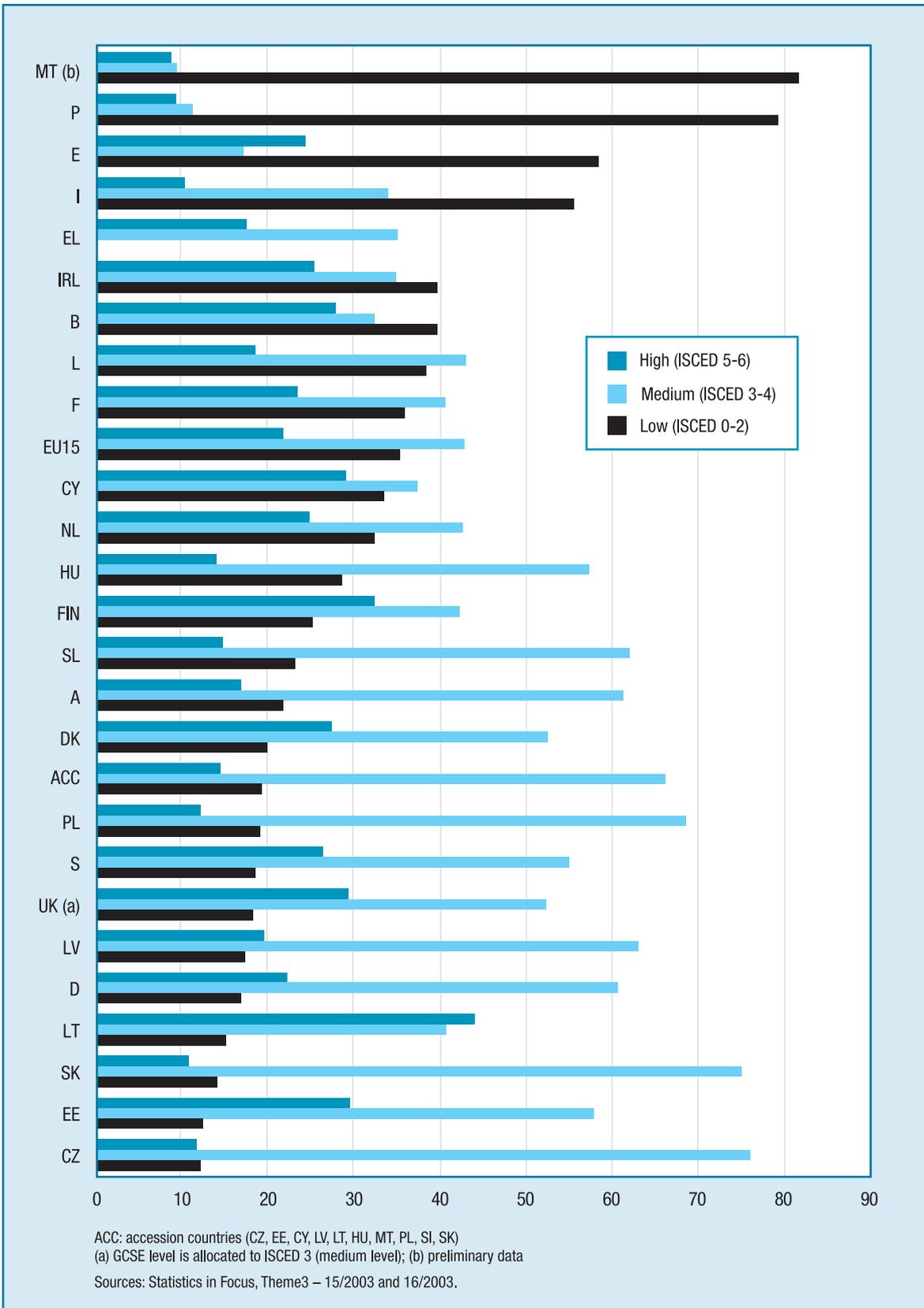
Compared to EU15, the working age population of accession countries is on average better educated (more people hold at least an upper secondary level diploma). However, this masks – as in the EU15 – great differences between countries (Figure 3.2, 18.3 % of the working age population in Malta and 87.8 % in the Czech Republic have attained ISCED 3 level at least).

It is a well-known phenomenon: the educational level of the European population is rising. Younger generations study longer and obtain higher credentials than their elders. The proportion of people completing upper secondary education (ISCED 3-4) has increased in all countries, including the accession countries (Table A5 of Annex 3). Overall in the EU15 and accession countries, the number of people with low level of education has decreased spectacularly between the two generations while the

⁽⁵⁾ Assuming a constant VET participation in each generation, the individual demand for VET would change only according to the size of the age cohorts. In this case, estimation of pressures or relief on VET capacities could be forecast relatively easily with demographic projections. VET policy would, therefore, be in a position to prepare in time for these fluctuations (Tessaring, 1998). However, the demand for – and participation in – VET among young people is not likely to be constant given the changes experienced both on the supply side (general increased attainment level, changes in the status of VET, etc.) and on the demand side (level and type of qualifications needed in the labour market, sectoral changes, etc.). Forecasts need, therefore, also to take into account change in the social demand for education.

⁽⁶⁾ In addition to early retirement, people tend to stay longer in education. Consequently, the population working and contributing to social security funds tends to become smaller compared to the proportion of the population depending on this system.

Figure 3.2. Education level of the population, aged 25-64 years, EU15 and accession countries, 2002 (%)



number of both upper secondary and of higher education graduates has increased (Figures 3.3 and 3.4).

Women, especially, are obtaining higher level diplomas. The proportion of low-qualified women (ISCED 0-2) has been halved in 30 years, while the number of those obtaining higher education diplomas (ISCED 5-6) doubled (Figure A2 of Annex 3).

The EU is trying to achieve a further reduction in the number of young people leaving education with a low level of qualification (i.e. without having reached at least the upper secondary level). Looking at recent trends, this number appears to be declining in almost all countries, though at a different speed and from different initial levels (Table A6 of Annex 3). It remains to be seen whether the target set under the Lisbon strategy of each country halving the number of non-qualified young people by 2010 ⁽⁷⁾ can be reached and how it will affect further increases in educational levels ⁽⁸⁾ (for more details on the European Benchmark for education and training systems see Box 3.6 and Section 3.3).

These trends may have the following consequences for education and training:

- (a) education and training pathways at lower secondary level are progressively changing from qualifying pathways to routes providing basic education targeted at pursuing studies at higher levels;
- (b) upper secondary pathways will have to become increasingly diversified (to cope with the variety of needs and wishes of participants) but also more polyvalent (allowing both qualification and continuing with studies in higher education). Making possible education and training routes more flexible is required;
- (c) safety net measures will have to be developed for those who drop out of schools early as the population of 'low'

achievers is at risk of being stigmatised and disadvantaged in a labour market where the majority of newcomers will have higher credentials;

- (d) CVT, including integration programmes and active labour-market policies, has an important role to play in avoiding exclusion and promoting the upskilling of low skilled workers already in the labour market (see below).

1.4. Skill renewal through adult learning and CVT

Participation in adult learning and CVT in enterprises is part of the skill supply in the labour market. Lifelong learning is now a framework guiding education and training policies and reforms in many European countries (Section 2.1.1 and Annex 4 presenting the main results of the OECD thematic review on adult learning).

Recent data shows that 7.0 % of people aged 25-64 in the EU15 and accession countries participated in adult learning ⁽⁹⁾. However, the figure varies greatly across EU countries, from 0.4 % in Greece to 20.4 % in the UK. In the accession countries, participation in adult learning is generally low, ranging from 2.4 % of people aged 25-64 in Estonia to 8.4 % in Slovakia (Table A7 of Annex 3). In this respect, one should note that many factors are conditioning the development of – and participation in – adult learning. One that has been identified is how well developed the initial VET system is. Initial training and CVT seem, in part, to be mutual substitutes: in countries with a strong and high level of IVT, CVT tends to be less developed (and vice versa; Descy and Tessaring, 2001a: 327). Figure 3.5 also shows that participation in continuing education and training

⁽⁷⁾ 'By 2010, Member States should at least halve the rate of early school leavers with reference to the rate recorded in the year 2000, in order to achieve an EU average of 10 % or less.' (EC, 2002b: 11).

⁽⁸⁾ In addition to the target of reducing the number of early school leavers, i.e. people with low levels of education, the Education Council also agreed on 5 May 2003 that 'By 2010, at least 85 % of 22 year olds in the EU should have completed upper secondary education'.

⁽⁹⁾ Labour force survey data 2002, participation in some form of continuing training (excluding if the purpose is initial education and training) during the four weeks preceding the survey.

Figure 3.3. Education level of the EU15 population by age groups, 2002 (%)

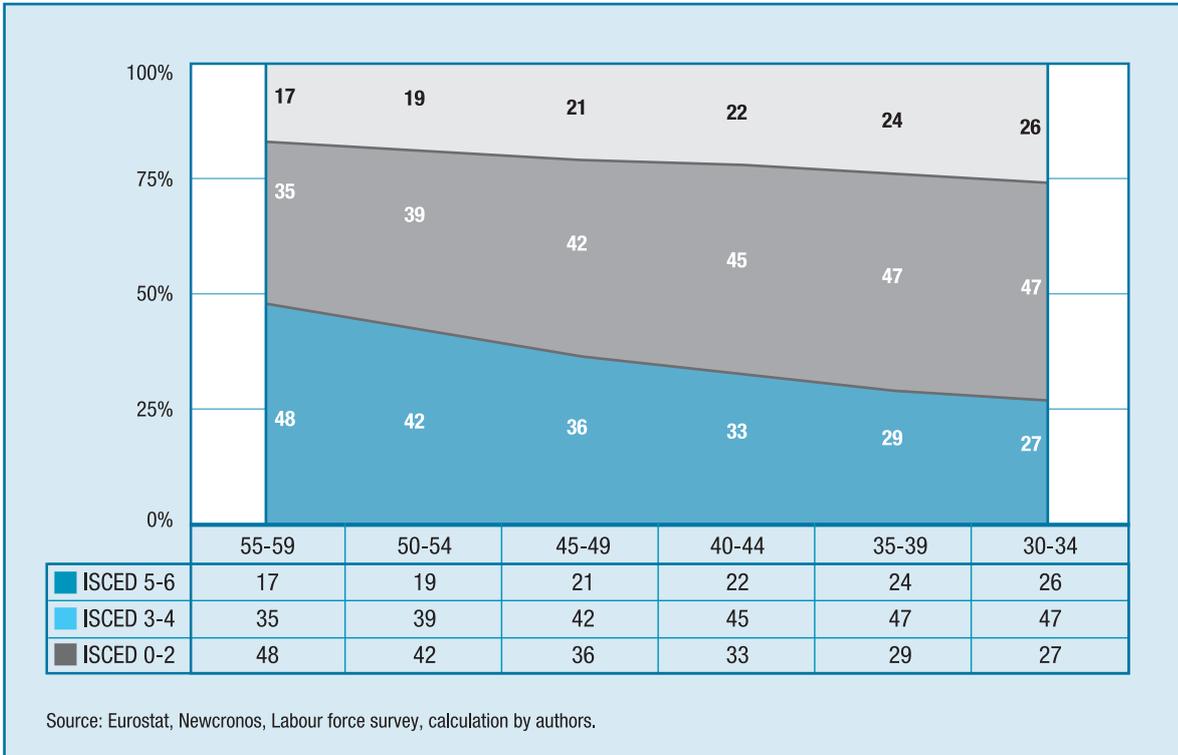


Figure 3.4. Education level of the accession countries population by age groups, 2002 (%)

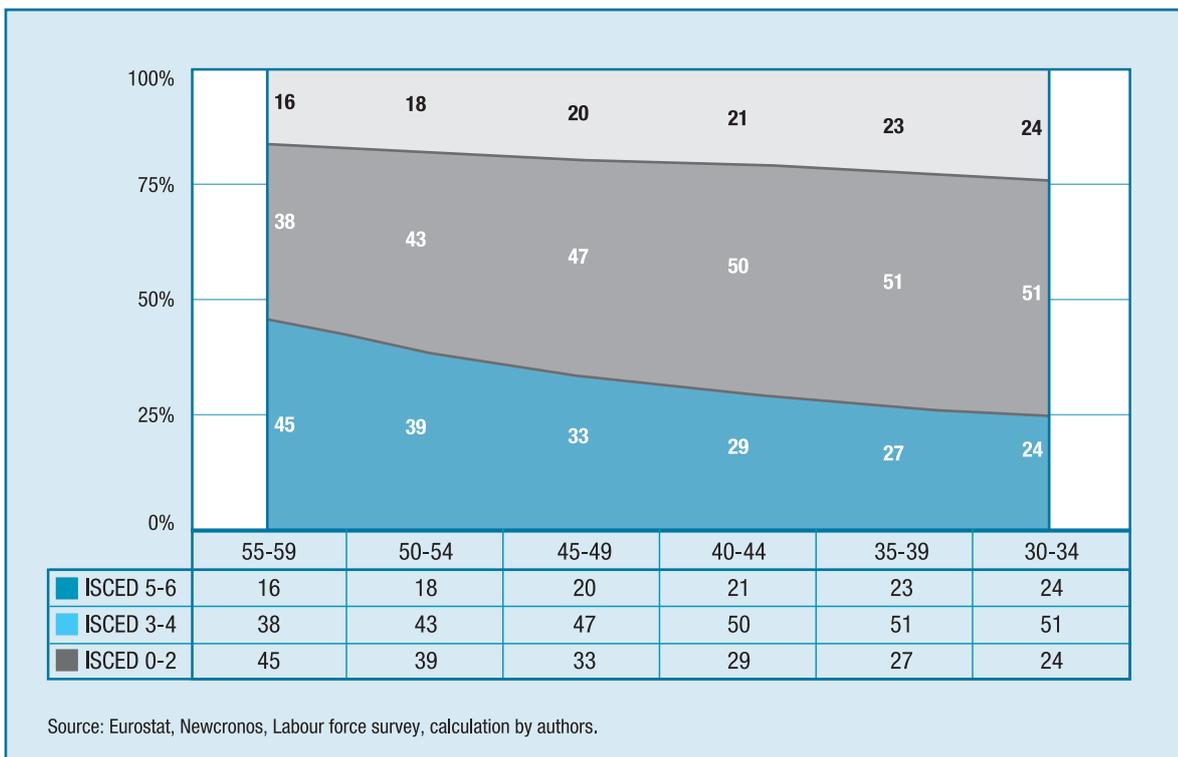
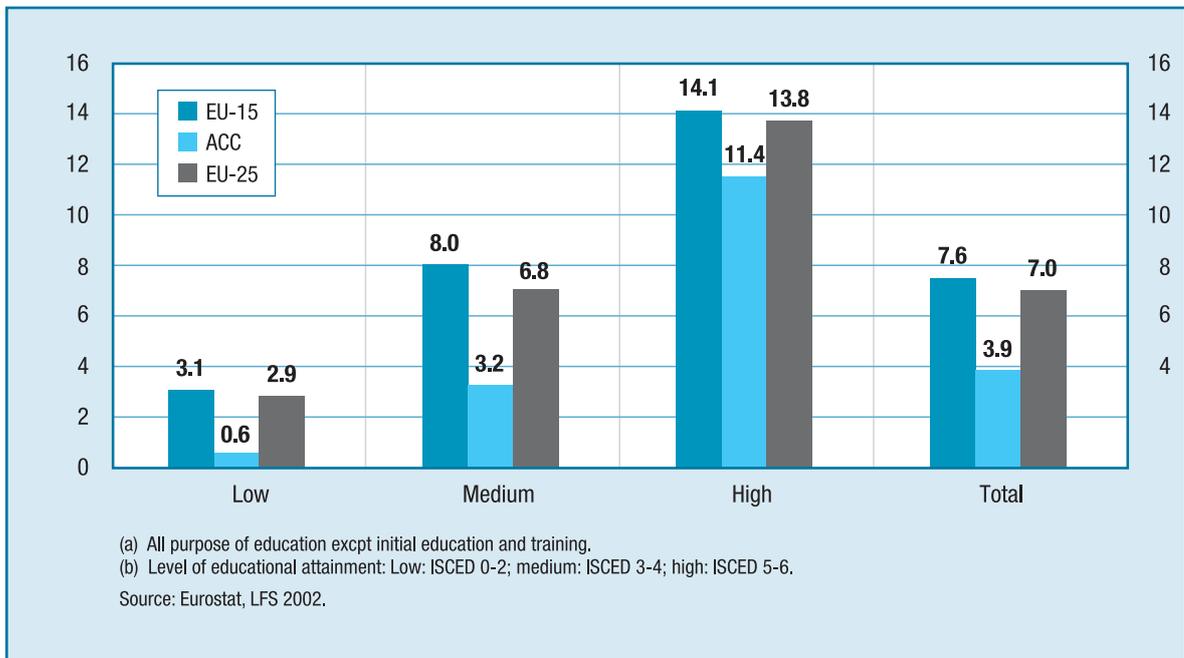


Figure 3.5. **Participation in continuing education and training ^(a) of the population aged 25-64 years by level of educational attainment, 2002, % ^(b)**



programmes is also clearly influenced by the initial level of education. The percentage of people who participate in continuing education and training clearly increases according to the level of educational attainment. Naturally, adult education and training is mainly undertaken to adapt to technological change, obtain promotion or upgrade skills (Table A7 of Annex 3).

At the same time, around half of EU15 citizens say they are prepared to pay partly or totally for the cost of training. However, this depends on the purpose of the training concerned; people are less ready to pay for job-related training. People seem more inclined to pay for training when personal returns are involved, for example when the purpose is to improve their private life, to learn a language or to obtain a recognised certificate (Cedefop, 2003; Table A8 of Annex 3) ⁽¹⁰⁾.

There seems also to be a growing awareness of the importance of the 'learning to learn' competence (which can be considered as an

indication of the awareness of the importance of learning throughout life). Younger age groups tend more often to find it very useful, both in private and working life, than older groups. There are also different perceptions between countries of the usefulness of 'learning to learn'. In Denmark, Greece, Iceland, Italy, Luxembourg, the Netherlands and Sweden, people are particularly aware of the usefulness of this competence in working life (over 90 % of people find it very useful) (Figures A3 and A4 of Annex 3).

The percentage of enterprises offering training varies very much between countries (from 18 % in Greece to 96 % in Denmark in 1999), with the EU15 average standing at 62 % (Figure 3.6). However, compared with 1993, it has been increasing in all countries (surveyed at that time) except in Germany. This could indicate a more voluntary training policy in enterprises, confirmed by the proportion of employees having participated in CVT courses provided by enterprises (which constitute only

⁽¹⁰⁾ The analyses presented here (as well as the data of Table A6, Figures A3 and A4) derive from the recent lifelong learning questionnaire included in standard Eurobarometer (wave 59.0) in 2003. More information on this initiative by Cedefop and the European Commission, Directorate General Education and Culture, is presented in Box 1.

Figure 3.6. **Training enterprises, EU15 and accession countries, 1993 and 1999**
(% of all enterprises)



Figure 3.7. **Participation in CVT courses, EU15 and accession countries, 1993 and 1999**
(% of employees, all enterprises)



Box 3.1. The 2003 Eurobarometer on lifelong learning

The questionnaire on lifelong learning designed by Cedefop and Directorate General Education and Culture was integrated into wave 59.0 of the Eurobarometer survey. It was carried out early in 2003 (January and February) in the 15 Member States and in Iceland and Norway .

Eurobarometer is a sample survey ^(a), representative of population aged 15 and more in terms of gender, age, region ^(b) and urbanisation size.

The questionnaire contained five modules which balanced the type of information collected:

- 1st module: background information on the individual and general opinion about relevant skills in different areas of life and about learning-conducive situations at work;
- 2nd module: past learning experiences and their benefits;
- 3rd module: learning preferences and reasons for learning in the future;
- 4th module: obstacles to learning, possible incentives and useful source of guidance on learning and career prospects;
- 5th module: people's opinion on lifelong learning and related issues (willingness to pay, informal competences, etc.).

(a) To establish the sample, the universe was stratified in terms of urbanisation level and region; on this basis, the target number of sampling points (i.e. the smallest administrative unit for which census data are available) is fixed for each cell of the matrix. In each country, a number of sampling points was selected at random, proportional to population size and to population density. In each of the selected sampling points, an initial address was drawn at random. Using standard random walk procedures, further addresses were selected as every Nth address from the initial address. In each household, one individual is selected at random. All interviews are carried out face to face in people's homes and in the appropriate national language. Up to four recalls were made.

(b) NUTS 2.

Source: Cedefop, 2003.

one form of possible CVT, Figure 3.7). The participation rates of employees in CVT courses has increased substantially in all countries for which comparison is possible, except Ireland and Greece (where the small difference observed might not reveal a real change in training policy of enterprises). It

must be noted that participation in CVT courses is much lower in accession countries than in the EU15, which suggests a tendency to organise more informal training in enterprises in these countries.

From these indicators, it appears that there might be a tendency for employers to provide more CVT and, therefore, increase learning opportunities for employees. Nevertheless, overall participation in adult learning is still very low in many countries, although awareness of its importance seems to be high. An increase in participation by adults in learning is a desirable trend for the development not only of economies but also of societies. It should not only be a government and public policy concern; employers, too, should become more aware of its importance. To increase participation of adults in learning is a driver for many reforms of education, training and lifelong learning policies and provision in Europe. However, the ways and means tend to differ greatly from one country to another.

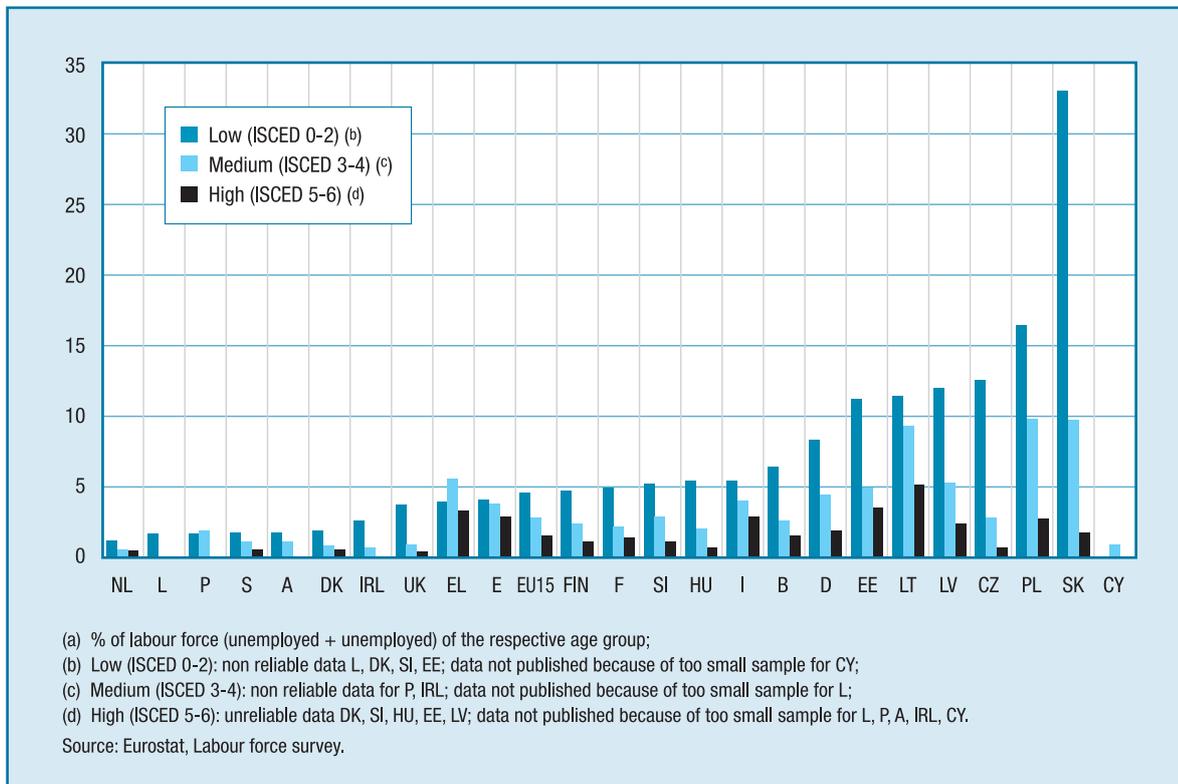
1.5. Educational profile and situation in the labour market

The higher the level of education, the higher is the employment rate. Those who study more are better off in the labour market. In the EU15 as a whole, those without qualification have an employment rate of 55 %; having achieved tertiary level qualification, this rate raises to 84.5 %. Accession countries follow the same pattern (Table A9 of Annex 3). Inactivity rates are also highly related to educational level and, at a same educational level, women are more inactive than men (Figure A5 and A6 of Annex 3).

In 2002, unemployment in the EU15 was 6.7 %, ranging from 2.2 % in Luxembourg to 9.7 % in Spain. In accession countries, unemployment rates were ranged between 2.9 % in Cyprus and 16.9 % in Poland. In most countries ⁽¹¹⁾, the higher the level of

⁽¹¹⁾ Except Denmark, Greece, Luxembourg and Portugal which present particular patterns.

Figure 3.8. Long-term unemployment rates ^(a), persons aged 25-64 years, by educational level, 2002 (%), countries sorted by ISCED 0-2



education, the lower the unemployment rate (Table A10 of Annex 3).

In addition, in all EU countries (EU15 plus accession countries) those with higher levels of education suffer less from long-term unemployment (LTU, 1 year and more, Figure 3.8). Lower-educated people everywhere (except Greece) are more affected by LTU. Compared with EU15 countries, the situation is worse in most of the accession countries.

Individuals are making a good choice when studying more, achieving higher level diplomas and undertaking training during working life because this increases employment opportunities and protection from unemployment. Governments are well advised in promoting and supporting these trends by various policies, programmes and measures, as better skilled workers participate more in

the production system and are more active citizens (Part 5). Is there, however, a real new demand in the labour market for these higher levels graduates and better skilled workers or are they simply ‘absorbed’ in the production system generating substitution and over-qualification? This will be examined in the next section.

1.6. Towards the knowledge economy: transformation of labour markets ⁽¹²⁾

Western economies have been undergoing structural change. This is the result of successive periods of crisis, globalisation of markets and increasing international competition, restructuring of economies towards services and changing work

⁽¹²⁾ This section is a brief summary of the main features of structural changes on the labour market and their consequences for skill requirement and renewal. For an extensive discussion of these issues, the reader should refer to Descy and Tessaring (2001a: 141-188, 227-299, 352-370).

organisation and production processes, accompanied by the constantly increasing pace of technological progress. All these factors influence the number and the type of jobs available as well as the skill profile required from workers (Descy and Tessaring, 2001a: 147-158).

The consequences of globalisation on work organisation, however, can mainly be observed in a limited number of big multinational or a minority of high-tech businesses. The majority of jobs can still be found either in SMEs (two third of the EU15 workforce, *ibid.*: 165-188) or in industries firmly-anchored to a given location (e.g. extraction and processing of raw material, building construction, transports, physical services, etc.).

Independently of globalisation, increased competition between enterprises, customisation of production and the massive introduction of ICTs require flexibility, innovation and reorganisation of the whole business and production processes. New performance criteria are imposed on companies, which try to reduce the division of labour to take advantage of human capital. This in turn influences the requirements for skills imposed on workers:

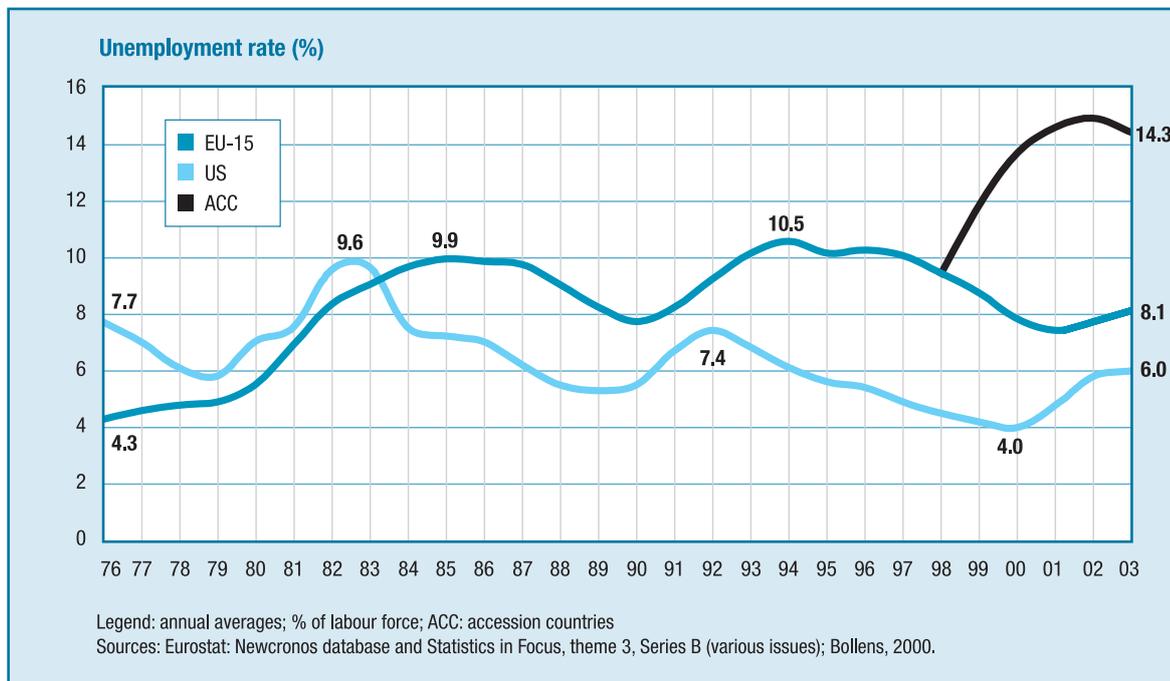
- (a) organisation and problem-solving competences are important to be able to plan, organise and control work autonomously or with colleagues in a team;
- (b) social skills and competences are required by increased team and group work as well as by more contacts with suppliers and customers;
- (c) work attitudes such as quality consciousness, reliability, accuracy, etc., are valuable in reducing costs and production time while keeping up with quality;
- (d) creativity and entrepreneurship are necessary to enhance the innovative capacities of enterprises;
- (e) international skills or intercultural competences are useful in dealing with foreign clients, suppliers, subsidiaries, etc.;
- (f) managerial skills are needed to deal with decentralisation of decision-making and of activities (*ibid.*: 155-156).

These skills requirements should be translated into objectives for (V)ET curricula. However, it would be illusory to believe that every job requires these competences to high degree. The polarisation of the labour market, with knowledge-intensive highly skilled jobs on the one hand and more repetitive and unskilled jobs on the other seem to increasingly characterise our economies (*ibid.*: 158-165).

This is one of the results of several crises since the early 1970s, which have made EU labour markets more precarious and polarised, characterised by generally high and persistent unemployment. To be in a relatively protected situation, individuals are required to be employable, adaptable and preferably (highly-) skilled. This does not mean that there is no longer a demand for lower qualified workers but they are more likely to be hired in the less stable segments of the labour market, occupying more precarious jobs, in worse-off industries or sectors (*ibid.*: 352-370).

As a consequence of these structural changes, transforming European economies into 'knowledge economies' or societies has become the political target of the EU (Lisbon strategy). This is seen as necessary to give Europe the world's most competitive economy. Giving access to, and increasing participation in, lifelong learning to improve workers' employability and adaptability has become an important objective. In addition, new challenges are posed for initial education and training. They no longer provide the qualification leading to a lifetime job. Instead they need to ensure the provision of foundation skills and competences on which individuals can then rely to develop their potential throughout their lives. Finally, both initial and continuing training (including active labour market measures) have a role to play in helping to avoid exclusion from the labour market of unskilled or low skilled workers (*ibid.*: 97-116).

Figure 3.9. Development of unemployment, EU15, ACC, US, 1976-2003 (%)



1.7. Employment and unemployment patterns in EU15 and accession countries

As a result of the developments described in the previous sections, qualitative and quantitative mismatches (between demand and supply) have appeared in the labour market. Unemployment is resulting from these mismatches.

Figure 3.9 illustrates the evolution and cyclical behaviour of unemployment in the EU15, in comparison to the US since 1976 and to the accession countries since 1998. In the early 1980s, unemployment in the EU surpassed that in the US. In the first half of the 1990s, EU unemployment reached an unprecedented level of 11 % but declined until 2001. In the accession countries, unemployment was much higher than in the EU15, with a peak of 14.8 % in 2003. Phases of increasing unemployment (1974-77, 1979-83, 1990-93 and after 2001) reflect major economic recessions due to the energy crises in the early 1970s and 1980s, the fall of the Berlin Wall and the end of the Soviet

Union in the early 1990s and, in the early 2000s, the crisis of the ‘new economy’ and the effects of September 11, 2001.

Recent indicators (2002) on employment and unemployment in the EU15 and accession countries, based on the Labour force survey data (Figure 3.10) allow further comparison between the two groups of countries.

- In the EU15, 160.8 million people and, in accession countries, 28.3 million had a job during the reference period for the survey (spring 2002). This corresponds to employment rates of 64.2 % and 56.2 % respectively:
- (a) 18 % of EU15 workers declared that they work part-time, of which about 14 % is involuntary part-time. In accession countries, part-time employment represents around only 7 % of total employment;
 - (b) in the EU15, 81 % of people working part-time are women, while in accession countries, women represent 61 % of this group;
 - (c) in the EU15 and in accession countries, 85 and 79 % respectively of people with a job were employees, 13 % of EU15 employees in the EU had a contract of limited duration compared to 11 % in accession countries;

Figure 3.10. Work status of persons aged 15-64, EU15 and accession countries (ACC) ^(a) (2002) ^(b)

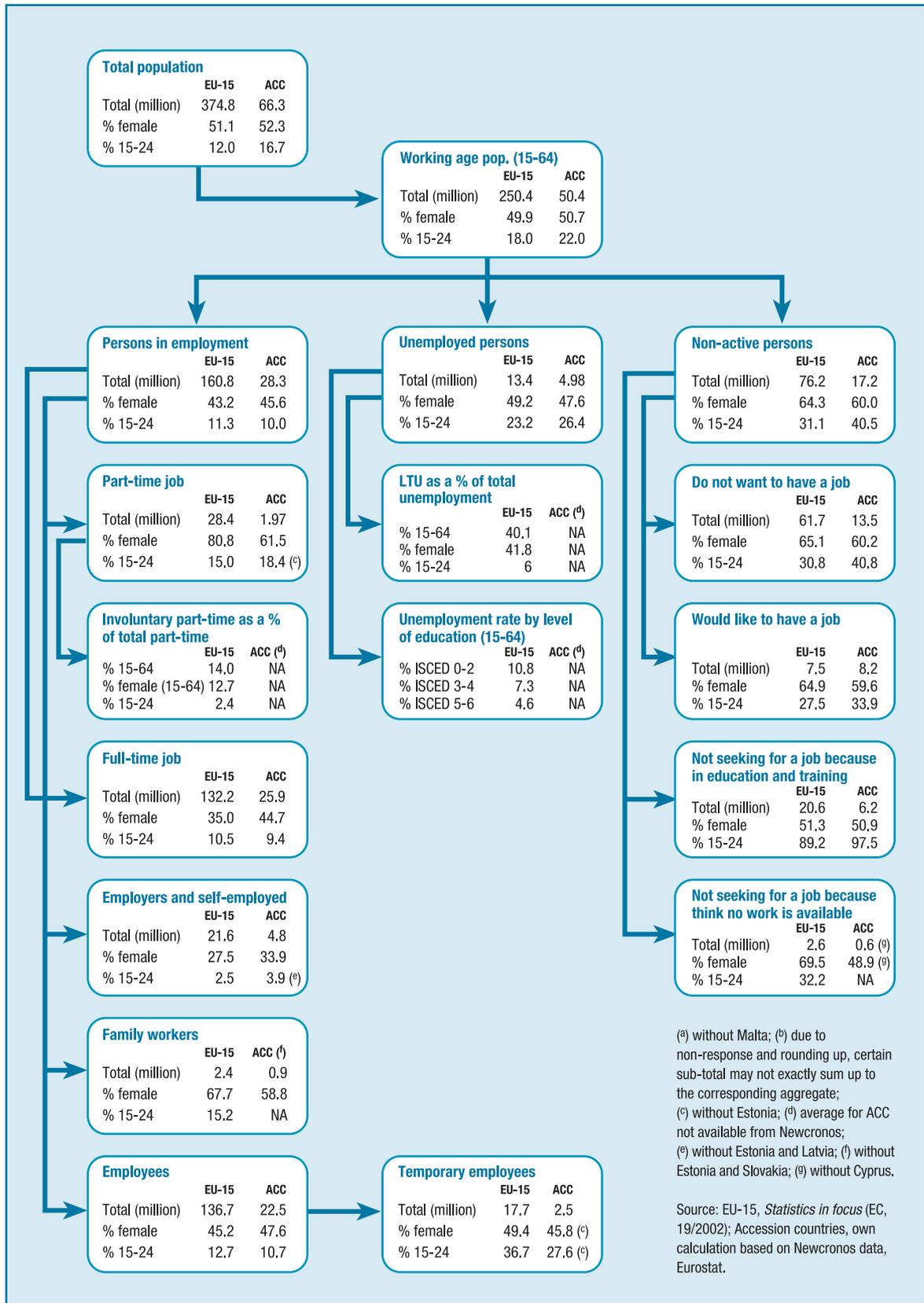
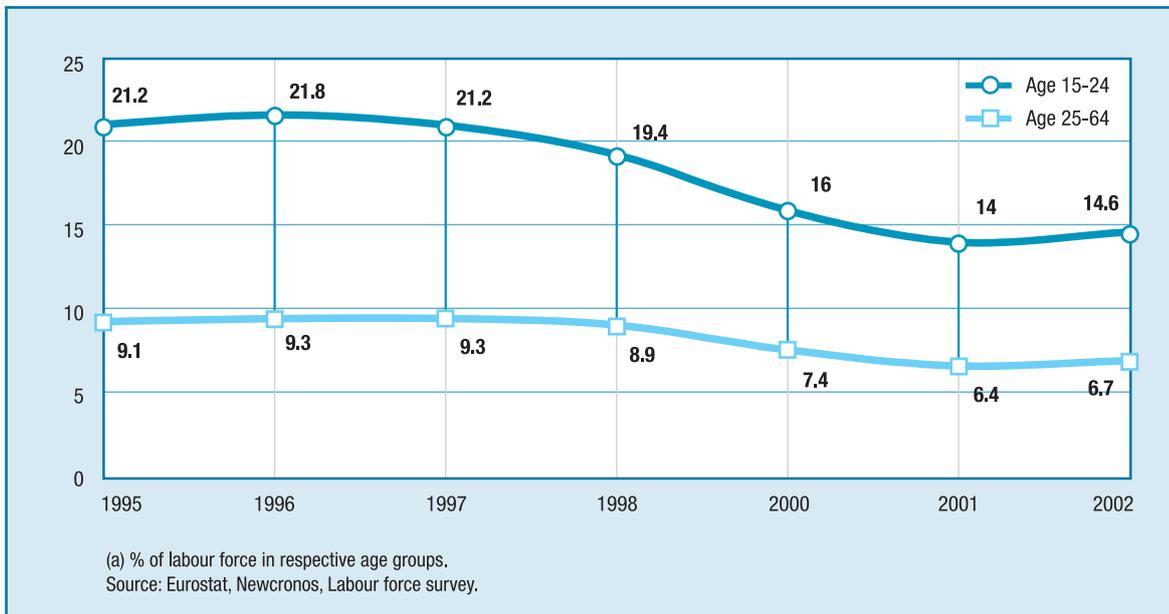


Figure 3.11. **Development of unemployment (a), comparison of young people and adults, EU15, 1995-2002 (%)**



(d) employers and the self-employed represent 13.4 % of total employment in the EU15 and 17 % in the accession countries.

In the EU15 13.7 million people (5.4 % of those aged 15-64) and in accession countries almost 5 million (9.9 % of those 15-64) were unemployed:

(a) LTU (more than one year) represents 40 % of total unemployment in the EU15 and in accession countries between 20.1 % (Cyprus) and 65.2 % (Slovakia);

(b) in the EU15, the unemployment rate of people aged 15-64 with a high level of education is 4.6 % compared to 10.8 % for those with a low level of education.

In the EU15, 76.2 million and, in accession countries, 17.2 million people aged 15-64 are inactive:

(a) 27 % of inactive people in the EU15 and 36 % in accession countries are in education and training; most of them aged 15-24;

(b) 10.1 million people in the EU15 and 8.8 million people in accession countries are

‘hidden unemployed’, i.e. they are inactive but declare they would like to work or are not seeking a job because they think no job is available.

Employment and unemployment patterns in the EU15 and in accession countries appear different ⁽¹³⁾; this has to be taken into account at many levels now that all these countries form one open market and, probably, on what should be the main axes of a common employment policy. Regarding unemployment, the situation of young people in the EU15 is much more difficult than for adults. Almost one in four (23.2 %) unemployed people are aged 15-24 (Figure 3.11). Unemployment among young people ⁽¹⁴⁾ in 2002 was more than double that of adults. There has been a clear improvement of the situation since 1995; unemployment decreased by almost 7 percentage points, from 21.2 % in 1995 to 14.6 % in 2002. In contrast, adult unemployment has not seen such a significant decrease during the same period (minus 2.4 percentage points).

⁽¹³⁾ These averages mask individual country variations and particularities, which in some cases can be large. This publication is not, however, the place to detail these changes.

⁽¹⁴⁾ 15-24 unemployed/15-24 in the labour force.

1.8. Conclusions

This brief review of changes in supply and demand helps to identify the following main drivers for reforming education and training and for introducing active labour-market policies:

- (a) the population of the EU is ageing and this trend will be accelerated by the accession of the 10 new countries;
- (b) the initial educational level of the EU population is improving;
- (c) skill renewal through lifelong learning is becoming increasingly desirable in EU countries;
- (d) better educated people are an asset for national economies not only because they participate more in, and yield benefits for, the production system but also because education generates considerable non-material benefits (increased social cohesion, better health, reduction of crime, etc., see Part 4, Chapter 4);
- (e) higher level skills are needed in modern global economies, especially in knowledge intensive-sectors, but at the same time labour markets are becoming more polarised, with more precarious and repetitive jobs for the less educated people;
- (f) unemployment will remain a concern for EU economies, especially in the context of enlargement; further effort will be required to fight unemployment through targeted policies and programmes.

These changing contexts require, on the one side, qualitative changes in education and training and, on the other side, measures aiming at the quick reintegration of people who are excluded (or at risks of exclusion) from the labour market. The following chapter will discuss reforms of education and training and VET systems and their evaluation, while Chapter 3 addresses active labour-market

measures and asks the question whether training works as a tool for reintegrating people into the labour market.

2. Principles, models and practices for evaluating education and training ⁽¹⁵⁾

2.1. Introduction

Reforming education and VET means taking actions that will change a part or element of the relevant system, influencing how institutions operate and interact. It involves changing the role of the different actors and, therefore, requires their acceptance and their understanding. It is process that has long-term implications and impacts. Additionally, reforming VET cannot/should not be undertaken in isolation from the other related systems (e.g. production).

Education and training system evaluation deals not only with assessing the results of reform, it can assist policy decision and accompany change. Traditional methods, aimed at assessing the effectiveness of a single programme, are ill-suited to understanding the complex set of interactions that regulate various institutions and actors in education and training; systems evaluation requires its own tools and methods.

In this chapter, we briefly describe some of the education and VET reforms currently undertaken in Europe. Then, we present what could be some principles and models for evaluating education and training systems and reforms. In Chapter 3, concrete examples are reviewed and discussed in the light of the principles and models discussed, presenting system evaluation and policy learning in action.

⁽¹⁵⁾ This chapter draws, among other sources, on the following contribution to the Background report: Coles (2004), *Evaluating the impact of VET reforms of vocational education and training: examples of practices*; Pont and Werquin (2004), *Look, listen and learn: an international evaluation of adult learning*; Viertel et al. (2004), *From project to policy evaluation in vocational education and training – possible concepts and tools*; Nieuwenhuis and Shapiro (2004), *Evaluating systems' change policies in VET – Learning from the Danish and Dutch cases*.

2.2. Renewing education and training policies

Governments can choose to design and prioritise various policies to improve responsiveness to new and changing demands in society and the economy, to enhance competitiveness, to improve social cohesion and to combat unemployment and exclusion from the labour market.

Beside education and training system reforms, targeted policies and measures (especially training), are implemented to combat unemployment and exclusion from the labour market more directly. Such active labour-market policies, and employment strategies and policies designed at EU level, are discussed in Chapter 3 together with evaluation results.

A medium and long-term approach is to design new education and training policies and to implement reforms in education and training systems. Many such discussions and reforms are current in Europe and are the subject of this section.

2.2.1. Coping with new contexts and challenges

The socioeconomic transformations that we described in the previous section set a new context to which education and training and VET systems must adapt to supply the type and level of skills needed by knowledge-based economies. A shift ought to take place from traditional education and training provision, focused on initial education and training, to developing CVT and learning opportunities throughout life. Education and training has to prepare individuals for the new, rapidly changing, and more precarious conditions they must cope with during their adult lives.

While carrying out a meta-analysis of reports discussing evaluations of VET system reforms, Coles identifies the following reasons for reforms: 'political wish or mandate to reorganise VET; economic necessity; relief of social issues; the need to find more equitable ways of funding VET; the need to overcome bureaucratic barriers; feedback from enterprises about skill needs

and shortages' (Coles, 2004: Section 4.3).

According to Coles again, employers have been motivated to participate in (reforms of) training activities for reasons such as 'improving the quality of training; staff replacement; company growth; availability of government funding for training; improving the general level of education of the workforce; developing a more motivated, confident and enthusiastic workforce; developing basic literacy and numeracy skills; improving the specialist skills base; improving communication' (Coles, 2004: Section 6.4).

Private initiatives and provision should be fostered and public provision of adult education and training increased (especially for people not in employment). Education and training are unlikely to receive major net funding increase as a whole, so budgets will have to be shifted and a greater efficiency achieved.

For public authorities this means:

- (a) adapting curricula and qualification levels provided by education and training to the type and levels of skills needed in the labour market in the short-, mid- and long-term;
- (b) making education and training systems more responsive to local and particular labour-market needs and adaptable to new and changing demands;
- (c) using education and training to avoid exclusion, especially long-term exclusion from the labour market, and to increase equity and general well-being in society;
- (d) supporting and creating opportunities for learning during the whole lifespan of individuals, including training at the workplace;
- (e) making education and training and VET more efficient, so providing the necessary skills and qualifications as broadly and inexpensively as possible.

For individuals it means:

- (a) acquiring, during initial education and training, the type and level of skills which will allow them to integrate quickly and successfully in the labour market;
- (b) continuing learning throughout their lives, using as foundation the skills acquired

- during the initial phase of training and being able to valorise those acquired later in life;
- (c) having access to learning opportunities during adult life, leading to personal and social development, adaptability and lifetime employability.

For employers it means:

- (a) demanding/ensuring that young people leaving education and training are both specialist (immediately operational) and generalist (quickly and easily adaptable in the short, medium- and long-term);
- (b) providing formal education and training opportunities to ensure skills updating and also generally increasing the motivation and commitment of staff;
- (c) developing a learning-conducive working environment so that employees have regular and meaningful opportunities for non-formal and informal learning at work.

According to Nieuwenhuis and Shapiro (2004, Section 1), all VET systems are seeking a new equilibrium:

- (a) between initial VET and lifelong learning: the old infrastructure built for initial VET, the new one able to support lifelong learning;
- (b) between traditional occupations and flexible qualifications. Although traditional occupational profiles are slowly disappearing or changing content, the institutions are still built on the traditional occupational structure;
- (c) between school-based learning and qualification through work experience;
- (d) between social demands and economic markets: should the student be adapted to the labour market or can the worker shape his own economic context?
- (e) between employment and entrepreneurship: how to develop more innovative and autonomous workers.

Discussing the results of the review on adult learning undertaken by the OECD (2003a), Pont and Werquin (2004, Sections 4.5 and 4.6) note that all countries involved in the review ⁽¹⁶⁾ are designing new policies or taking actions

to implement reforms in adult and lifelong learning (see selected results in Annex 2). They also identify a high degree of convergence between the themes treated in the different country reports, revealing that countries are facing the same challenges in the field of adult learning. They underline that there is no better (or best) model on the practical modalities of adult learning when it comes to objectives such as efficiency or equity. Common aims may lead to different objectives and/or priorities. Systems are immersed in different political and social structures and involve the various key players in a specific manner.

Although specific recommendations have been made by the OECD for each country as a result of the thematic review, some recommendations are common to most, if not all, countries reviewed. They concern issues such as:

- (a) the need to adopt better concerted approaches;
- (b) recognising non-formal and informal learning;
- (c) developing financial incentives to support adult learning;
- (d) the need for more systematic evaluation;
- (e) the need for more and better guidance and counselling;
- (f) the diversification of the provisions, making them more attractive and responsive to demand (Table 3.1) ⁽¹⁷⁾.

In the transition countries (Viertel et al., 2004, Section 4.3), transformation of the socioeconomic context has been even more dramatic and the whole logic of VET systems has changed. Economies under communist regimes were characterised by a close relationship between the human resource needs of the economy, companies and vocational schools. With the change to a market economy, this no longer applies. Consequently most transition countries have put emphasis on increasing the responsiveness of VET systems. It is seen as crucial for the constantly changing and largely unpredictable requirements of labour markets for skilled

⁽¹⁶⁾ Covering Canada, Denmark, Finland, Norway, Portugal, Spain, Sweden, Switzerland and the UK.

⁽¹⁷⁾ For a detailed discussion of the OECD policy review method of evaluation, see Section 3.2.

Table 3.1. Recommendations made to countries as a result of the OECD thematic reviews on adult learning

Country recommendations	CA	DK	E	FIN	NO	P	S	CH
Develop a coordinated approach	X	X	X	X	X	X	X	X
Strengthen evaluation	X	X	X	X	X	X	X	X
Recognise informal learning	X	X	X	*	X	X	X	X
Introduce financial incentives	*	X	X	X	X	X	X	X
More outreach, information, guidance and counselling	X	X	X	X	X		*	X
Focus on specific groups: women, SMEs, unskilled, disadvantaged groups, etc.	X		X	X	X	X	X	X
Set up or make better use of discussion forums	X			X		X	X	X
Develop partnerships	X	X	X			X	X	
Devise tailor-made courses			X	X		X	X	X
Develop the use of information and communication technologies (ICTs) for distance learning			X	X				X
Step up and develop workplace learning		X	X	X	X			
Improve quality assurance strategies	X	X						
Introduce a competence-based system				X	X			
Assist with regional disparities	X		X	X				

* In some countries, recommendations were not made because mechanisms or policies were already in place or well advanced. This was the case, for instance, with the recognition of experience in Finland which has been in place since 1994.

Source: Pont and Werquin (2004, Table 3.2).

labour and also the growing demands of young people for higher levels of education and for more diversified and individualised learning processes and pathways. One can easily see that, although the needs might be more pressing in transition countries, their challenges are in many cases similar to those faced by West European economies.

2.2.2. Types of reform

The types of education and training- and VET-related reforms attempted in many European countries (sometimes in isolation, sometimes in combination) can be summarised as follows ⁽¹⁸⁾:

- (a) increasing parity of esteem between general education and VET, bringing curricula closer and providing access to
- (b) higher education from upper secondary level VET programmes;
- (c) developing new pedagogies and methods, more learner-centred, while renewing the curriculum's objectives and content;
- (d) increasing the emphasis on work experience during education and training and on work-based learning;
- (e) introducing new forms of assessment and validation of competences, including non-formal and informal learning;
- (f) introducing new qualifications frameworks or changing the qualification structure through the introduction of modular systems;
- (g) diversifying education and training provision by promoting new institutions and training providers, often accompanied

⁽¹⁸⁾ The short listing of reforms below does not attempt to be exhaustive. We do not discuss here the various reforms in detail but focus on how the reforming process can be steered, using evaluation, and how results and impact of reforms can be evaluated. This will be the subject of the next sections. The reader interested in detailed reviews of various reforms of education and training systems in Europe may refer to Descy and Tessaring (2001a, various chapters).

- by decentralisation of provisions and of management of institutions;
- (g) changing funding mechanisms and criteria to attain efficiency, while preserving equity and quality;
 - (h) involving the social partners (employers and trade unions) more in education and training decision-making, education planning and provision;
 - (i) reorganising and creating authorities and bodies responsible for adult learning and developing a training market where public and private institutions compete.

In the context of transition countries, Grootings (1993, cited by Viertel et al., 2004) describes different stages of reform ⁽¹⁹⁾:

- (a) 'corrective reforms that are initiated with immediate repairing objectives in countries emerging from a war or deep economic recession period;
- (b) modernising reforms aimed at reducing gaps and catching up with Western institutions. They are especially active at the level of curricula, teaching and learning methods, examinations, and schools textbooks;
- (c) structural reforms that are targeted at the structures, legal framework and management of educational systems;
- (d) systemic reforms that are deeper and have a global character because they call for a genuine change of paradigm in terms of educational policy. They are aimed not only at the curricula or the legislative framework, but also at the very internal logic of education and its relationships with the global social system. A systemic reform examines the key elements of every educational policy: the role of the state, the relations with the labour market, the financing system, the efficiency control, the normative role of national standards, etc.'

These stages of reform could also be applied as a grid to describe and categorise reforms implemented in western European countries. Corrective reforms would include labour market and social programmes and measures, aimed, for example, at the quick

reintegration of unemployed or other excluded categories into the labour market. Modernising reforms would include the isolated efforts of, for example, modernising curricula or developing new forms of assessment. Structural reforms could be decentralisation or renewal of steering and financing mechanisms. Finally, systemic reforms, which are probably more scarce, could encompass (some of) these different aspects in an integrated way (e.g. the implementation of a coherent and comprehensive lifelong learning strategy).

It appears that the socioeconomic factors identified in Chapter 1 are quite common to all European countries and they lead to similar challenges. All the developments and new policies concerning education and training are assumed to contribute to increasing the responsiveness of VET systems, individuals' adaptability and adequacy of skills for labour-market needs.

Parkes et al. (1999: 27, cited in Viertel et al., 2004, Section 2.2) suggests that all successful VET systems, independently of their cultural and social contexts, should possess the following elements:

- (a) 'to be able to define occupational sector priorities (on the best possible evidence base);
- (b) to be able to identify appropriate occupational sector competences and skills required [...];
- (c) to be able to turn these into curricula profiles and programmes and measurable standards;
- (d) to deliver these at school level [...];
- (e) to help make the processes attractive to students and teachers (transferability, visibility and portability of qualifications for students and working conditions for teachers);
- (f) to provide timely and effective feedback through evaluation, monitoring, quality control and tracer [cohort and biographical] studies for school leavers'.

Because of particular national contexts, due to historical, cultural and systemic differences

⁽¹⁹⁾ The first stage was proposed in 1997 by Birzea (1997).

as well as of different policy priorities, solutions, reforms or priorities may differ. There is no blueprint for a ‘good model’ or ‘one best way’. The transferability of policies and practices that are successful in one country is also questionable. Each country should follow its own path, in relation to its own context.

‘EU experience has shown that neither the common factors pressing on national policy formation for education and training, nor the priorities that most EU Member States share, have led to a uniform pattern of convergence (even if some trans-European trends have emerged).’ (Viertel et al., 2004, Section 6.2.5)

2.2.3. Conclusions

European economies are facing common challenges, driving them to reform education and training and VET systems and to develop active labour-market policies. The common aim is to address transformation of the economies, labour markets and societies that are resulting in changes in labour supply and demand.

As a result, the responsiveness and flexibility of education and training systems must increase to provide both individuals and the labour market with the appropriate skills and competences. Because of different political contexts, historical, cultural and systemic difference, attaining this goal will, however, take a different form at a national level and there is no blueprint of a good model.

‘Researching these system changes can be done best from an evaluating perspective. [...] [Implementing change in] VET systems can be seen as a kind of governmental learning, depending on the specific problem definition and the specific configuration of institutional and organisational actors and their stakes, policy strategies [and objectives] [...]. There is no one right way [...], although a European approach suggests a convergence of targets and goals. [...] Examples from other countries can be used as good practices, but should be adapted to one’s own national situation.’ (Nieuwenhuis and Shapiro, 2004: Chapter 1).

Due to both the common transformations observed and the laboratory of reforms and new policies generated, there is still a real

need for policy learning both within and between countries. Mutual learning (between countries and between stakeholders in one country) should take place and debates should be triggered by investigating others’ ‘solutions’. This is one of the ways that leads to the development of learning economies. Evaluation in its different forms can contribute to this process.

Results and impacts of the various reforms of education and training undertaken should be explored. However, ‘[...] policy evaluation [...] [is a] complex terrain. Concepts and methods are derived from sociology, economics, psychology, history, pedagogy, philosophy, management and organisational theory, comparative education, [...] etc., whose scientific theoretical position and methods have to be integrated’ (Viertel et al., 2004, Section 2). This is especially true in education and training policies and reforms. Methods and tools for policy evaluation are often implicit rather than explicit and transparency of concepts and methods used in policy evaluation is limited. There is a particular need for systemic evaluation methods.

2.3. Systemic analysis: a key to policy evaluation in VET

Education and training covers wide-ranging activities from initial mainstream education to short-term, highly focused, job specific skill development and much informal learning via socialisation and participation in communities of practice. It is located at the intersection between two basic human activities: learning and working. There are also multiple layers of activities in terms of where the learning takes place, who organises it, who pays for it and who certifies it. (Coles, 2004: Introduction)

Therefore, education and training and VET are very complex social systems and changing them takes a long time. VET reforms require interactions, debates and compromises between the main institutions and actors in learning and working systems, which do not share the same priorities and prime objectives.

They are also susceptible of having different perceptions of current problems and future needs. Some (re)conciliation needs to take place within VET.

In respect to methods for evaluating VET systems, Viertel et al. (2004) note that while a fairly robust methodology for project or programme evaluation exists, this is not the case for system evaluation. There is a need for tools, methods and theories which can be used for supporting policy design and implementation but also to evaluate policy performance when it comes to systemic change.

Von Bertalanffy, in 1950 (cited by Viertel et al., 2004: Section 4.1) defined systems as follows:

‘A system is a set of units with relationships among them [...] the state of each unit is constrained by, conditioned by, or dependent on the state of other units. The units are coupled. Moreover, the system as a whole has “got something” which its components separately have not got. [...]’.

Being a social system, VET depends on self-regulating mechanisms to maintain its boundaries and its continued existence within these boundaries. It is also in permanent interaction and feedback with other social systems, such as production. Evaluation of VET, therefore, requires systemic approaches focusing on the ‘analysis of relationships, communication channels, and responsiveness and adaptability, based on the fundamental understanding that changes in one component lead to changes in other components and in the system as a whole’ (Viertel et al., 2004, Section 4.2).

Policies change systems by influencing elements of these systems. Designing and evaluating policies means understanding how a particular element contributes to the totality (Viertel et al., 2004, Section 3.1). This hermeneutic circle has to be part of designing and implementing policies and reforms as well as evaluating their success.

Designing, implementing or evaluating education and training and VET policies is only possible through a systemic approach, applying, in particular, concepts of internal or

external consistency. From these, one can define where to intervene, how best to implement an intervention or how to evaluate the success of a policy.

Changing complex social systems needs coherent, persistent and consistent political actions at all levels of the system. Viertel et al. (2004, Section 4) insist on the need for systemic approaches to reforms by discussing the lack of internal and external consistency in improving education and training and VET.

2.3.1. Internal consistency

Internal consistency means that different elements of the system work together. A policy is due to fail if it focuses only on reforming one element without paying attention as to how it interacts with and influences the others, (e.g. reforming curriculum without taking into account teacher training). Strategies of reform often focus on one element of education and training at a time, failing to achieve consistency between learning opportunities and pathways. For example, one of the reasons why young people prefer academic stream education to VET is the wider possibilities they provide in terms of further studies. Therefore, a generally adopted solution is to open access to universities at the end of VET pathways. This is due to fail if the logic and selection criteria with which universities operate are not taken into account. Improving VET thus means implementing a systemic approach encompassing all the components of the system (Viertel et al., 2004, Section 4.3.1) (see Box 3.2).

2.3.2. External consistency

External consistency concerns the interactions of education and training with other social systems, especially the labour market. Changes cannot be implemented in VET systems without taking into account the needs of the production system and the consequences the reforms of VET will have on skill provision. External consistency is attained by articulating and regulating the tensions between the pedagogical logic of the schools and the competence needs of employment (Viertel et al., 2004, Section 4.3.2) (see Box 3.2).

Box 3.2. Lack of internal and external consistency while implementing reforms of VET in selected PHARE countries

One of the main objectives of the PHARE programme in Central and Eastern European countries is to support the process of reforming VET to make it more responsive to the demands of the market economy. VET has to be made more flexible to respond to changing demands as well as to individual needs and preferences. This should ideally be achieved while ensuring a certain consistency of VET provision and qualifications, thus avoiding a lack of national standards and fragmentation of the system.

In Bosnia-Herzegovina, the reform of VET included changes in VET equipment, curricula and methodology. However, vocational teacher training was not originally a project component. The reform was therefore not fed into pre-service (and in-service) training of teachers. Inspector training was neglected as well. This created a serious barrier to a broad acceptance of the new policy. This situation was corrected later and a budget reserve of EUR 200 000 allocated to run a programme for teacher and management training.

The curriculum design adopted in Estonia, modular-based and employer-led, is changing the logic of the VET system. Skill needs and educational goals have been identified and occupational profiles translated into curricula, while employers and their representatives are influencing VET provision. However, the infrastructure supposed to support curriculum implementation is very fragile. Clearly the National Centre for Examination and Qualifications is methodologically and technically too weak. This Centre, whose role is to award and accredit qualification, is the basic cornerstone of the new VET system. This puts the momentum of the innovation process and the fragile interaction between the employment and the education system at risk.

In Romania, a pilot Phare VET reform was implemented in about 10 % of the VET schools. After that phase the government decided to carry out this reform in all schools. However, no new equipment or training was provided to accompany this. The national inspectorate, only half familiar with the reform, was asked to support the process. The careful necessary planning and the substantial resources needed to procure new equipment and to train both teachers and school managers were not present.

Source: Viertel et al., 2004, Section 4.4.

Similarly, the US Department of Education stated in 1975: 'Systemic reforms address all the mutually reinforcing structures, processes and activities within the educational system, recognising that altering any one part of the system necessarily impacts on all other parts'.

In the 1990s there was growing interest in understanding the driving forces behind systems changes and how innovations in VET can be spurred through educational research in economies, where learning and knowledge is generally agreed to be of growing importance (Nieuwenhuis and Shapiro, 2004: Chapter 5).

In programme evaluation, there is also a trend to broaden the view from the focused and traditional evaluation of a programme, organised as if it was an independent entity to a more systemic policy view. In this respect, Grubb and Ryan (1999: 109) comment: '[...] a programme is conceived of something that can be created relatively quickly, introduced among the other institutions of a society, evaluated as a discrete entity, expanded or contracted. We call this a "project" or "programme view", because of its tendency to think of a VET programme as self-contained and independent. Then, what is conventionally called "programme evaluation" assesses the effect of the programme only, independently of any surrounding policies and institutions. We contrast this with what we call a "systems" view.' They elaborate later on (ibid.: 116): '[...] only rarely does the evaluation of VET programmes conceive of a larger system of programmes [...] a system that might develop slowly over time. [...] A corollary is that more energy is put into developing new programmes and evaluating them – and then abandoning them or trying new approaches – rather than continuing to develop institutions over longer periods of time [...]. This suggests that the entire evaluation enterprise, in its "programme" focus, is part of an incoherent and fragmented approach that is unlikely to lead to more effective VET policies over time. A system's perspective would, on the other hand, encourage thinking not about individual projects, but about widely available programmes that are linked to one another

and institutionalised'. We will elaborate further on how to implement this view in the part dedicated to programme evaluation (Chapter 4).

The basic argument is that any policy (whether a focused intervention or programme or large-scale reform) will interact with the rest of the system (institution, organisations, actors) and with other systems (such as production). This has to be taken into account both in the design of the policy itself and also of its evaluation.

2.4. Evaluation, at the service of policy-making in education and training

In the following sections, we will discuss some basic principles and models that apply to the evaluation of a system's reform. First, we will analyse how evaluation can assist decision- and policy-making as well as support the planning of change. Then we will look at different summative options that can be applied in the context of evaluating strategies of reforms and education and training systems.

The principles and models to be discussed are derived from evaluation approaches applied mostly in Europe (some concrete applications and cases will be presented in Chapter 3). Some methods and techniques can be derived from programme evaluation, but the evaluation of a system's change requires its own new tools, methods and understandings as well as a more constructivist approach to, and use of, evaluation ⁽²⁰⁾.

When evaluating systems and implementing strategies of reform, one could try to distinguish between a formative evaluation, which seeks to influence and promote change, and a summative evaluation, which seeks to assess and judge results ⁽²¹⁾. However, in the context of system reform, it appears that

evaluation is rarely singularly formative or singularly summative. Strictly summative evaluations are difficult to carry out as institutional and system change develops over time and cannot be considered completely implemented within a finite and relatively short period of time. Although, some techniques can – and should – be applied to judge the achievement of some objectives of a reform, our argument is that systems are dynamic. Implementing change is a long-term process involving various actors, stakeholders and institutions, in which case formative and constructivist evaluation design are also indispensable.

Kelleher (2001, cited in Nieuwenhuis and Shapiro, 2004: Chapter 5) suggests different functions of evaluation ⁽²²⁾ (within an organisational context). The two first functions are formative while the third is more summative. This reflects the way we will attempt to discuss evaluation in the following sections:

- (a) evaluation as a 'communicative domain' in which debates, persuasiveness of arguments, weight of opinions and conflicts will contribute to social construction of policies and practices;
- (b) evaluation as a 'stimulus' for the creative minds of policy makers (and practitioners);
- (c) evaluation as a feedback mechanism on the effectiveness of and efficiency of policy (and practice).

The policy cycle is traditionally seen in different distinct stages: decision-making, implementation and evaluation. This traditional view tends to refer to phases at which evaluation could be applied to support policy: before, during and after. However, the separation of three phases of reforms, and therefore of different evaluation functions, is artificial. One should see these three stages as a cycle where each level interacts with and depends on the others. The initial phase of policy design and planning is a phase of analysis and summative evaluation of previous

⁽²⁰⁾ System and programme evaluation are not substitute but complementary.

⁽²¹⁾ For a more in-depth discussion of summative and formative evaluation concepts and philosophies, please refer to Part 1.

⁽²²⁾ Section 1.3 in Part 1 discusses in more detail the various functions that can be endorsed by evaluation.

and current policies, of their internal and external (in)consistency (see discussion on systemic evaluation, Section 2.3). This initial phase in turn influences greatly what happens in implementation, when choices will become practices, when evaluation can be used, as a constructivist learning tool supporting the implementation of change. Finally, assessment, highly dependent on the others, cannot be studied in isolation from them and ultimately constitutes the first stage of a new policy. Following this logic, evaluation in itself is a continuous cycle which aims at improving policy-making, anchoring policy choices in social contexts and systems, ultimately leading to policy, institutional, organisational and individual learning.

Coles also touches upon the different functions of evaluation in the context of reforming VET systems: 'Some represent a feedback system to central planners on progress and are mechanical in nature using targets or performance indicators as accountability tools. [...] Other evaluation processes are embedded in the reform process itself and offer feedback to a range of interested parties. Indicators and performance measures are less prominent here and are used to improve organisational learning. The evaluation is seen as a way of communicating thoughts and actions across a range of actors.' (Coles, 2004: Section 1.1).

This section will discuss possible approaches to, and use of, evaluation in the context of reforming education and training and VET in particular:

- (a) how can evaluation be used as a steering and support device, to accompany the implementation of change;
- (b) how can evaluation support policy decisions, the establishment of priorities, the choice of objectives, the means and level(s) of action and planning of implementation;
- (c) how can evaluation be used to assess and judge the results of a reform, how can it help to assess where a system or a country stands; how can it feed into the policy

cycle and contribute to policy learning.

Evaluation (and research) ⁽²³⁾, including programme and policy evaluation, is expected to contribute towards deeper knowledge and insights on which policy-makers can draw and act (Nieuwenhuis and Shapiro, 2004: Chapter 5). It has, therefore, to consider the different questions outlined above.

In Sections 3.1 to 3.3, we discuss different evaluation tools that can support policy choices and reform planning and the implementation of change. In Section 3.4, we discuss methods that can be used when evaluating the results and assessing education and training and VET systems and strategies of reform.

2.4.1. Understanding and implementing change: institutional and organisational reforms

Changing VET is a complex, multi-layered process that must comprise all system levels:

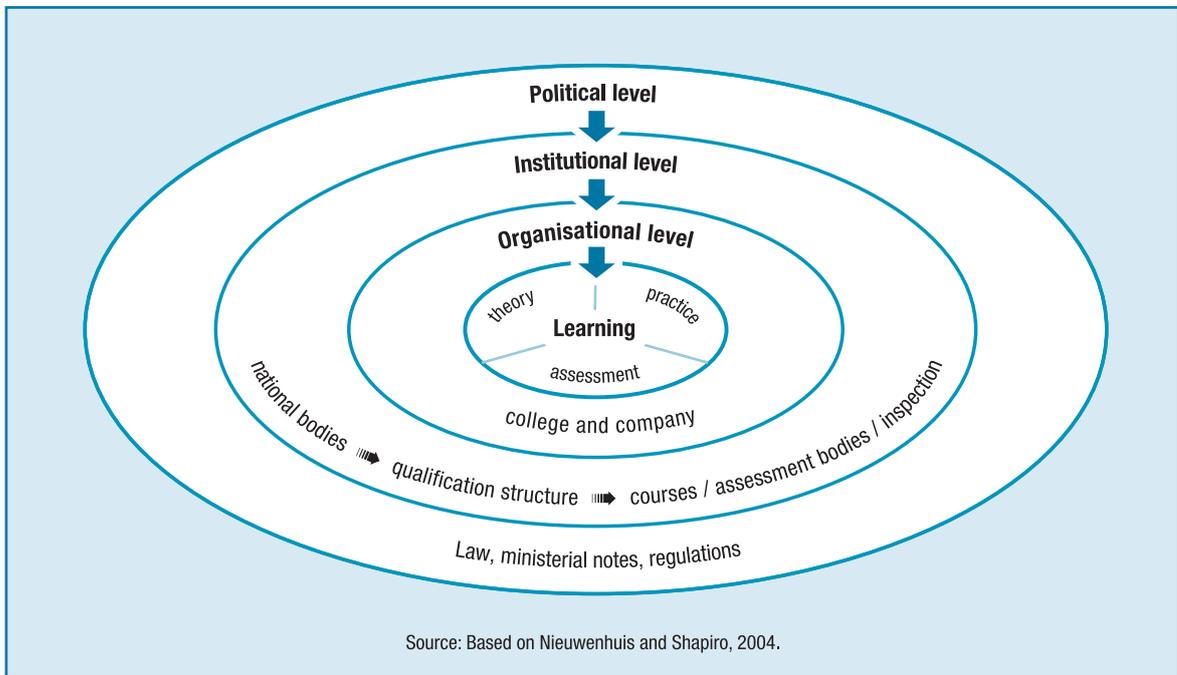
- (a) the primary learning process;
- (b) the organisational level;
- (c) the institutional frame, including the financing mechanisms;
- (d) the policy and legal frame;
- (e) the links to the other systems, such as employment and work (Viertel et al., 2004, Section 5.1).

Nieuwenhuis and Shapiro (2004: Figure 1) represent as follows the relationship between these different system layers.

In each country, VET institution and organisations have developed over time. At political level, governments, businesses and industries, unions, etc., have progressively built an institutional frame for VET (laws on education and labour, public-private funding arrangements, collective labour agreements, qualifications and wage structures, etc.). These institutions are difficult, or even sometimes obstacles, to change. In theory, organisations are easier to change than institutions but because organisations are dependent on institutions, organisational change has to be compatible with the institutional set up. Changing VET itself, the

⁽²³⁾ The differences and similarities between evaluation and research are addressed in the introduction.

Figure 3.12. Systemic levels within VET



heart of the system, has to be organised simultaneously through all the levels involved.

'Planning system change would be the art (rather than the science) of designing a multi-purpose, multi-actor, multi-stage process comprising coherent actions across many interdependent areas. This process must be based on a sound understanding of the institutional logic of the system and a number of hypotheses about the reasons for lack of internal and/or external consistency and how they could be overcome.' (Viertel et al., 2004, Section 5.4).

Reforming is determined by the institutional context of the country, the organisational settings and arrangements as well as the social values which all have grown historically. To understand and analyse reforms the evaluator should learn how a policy fails or succeeds in taking into account this complex framework. (ibid., Section 5.2).

VET system change is also complex because:

(a) it is a continuing process of incremental development, that does not always proceed smoothly one step after the other but can sometimes be combined with 'turbulence';

(b) it is often irreversible because system change has evolutionary characteristics. Once a process has started it cannot just be stopped and the initial situation retrieved;

(c) VET is related to production, education and social systems (Nieuwenhuis and Shapiro, 2004).

Changing VET is not only a concern for education, but also for the socioeconomic system and of cultural and social traditions. Policies for change cannot be organised top-down. Changing VET has to be directed and organised simultaneously at all levels. Interactivity and consistency between the different layers of the system (see Figure 3.12) are major requirements for effective change. 'Policy, intermediary structures, colleges, companies and teachers should interact in the change process; managing this process is like directing a large orchestra: if one part is out of tune, the whole performance is endangered.' (ibid. Chapter 1).

Accomplishing change is also about preventing or reducing resistance to change. The evaluator should, therefore, put in place a process to gain the support and understanding required from the different actors.

The constructivist thinking and the formative kind of evaluation deriving from it ⁽²⁴⁾, are especially helpful in a context where many stakeholders need to reach consensus. Here the views of stakeholders are considered as determinants and shapers of possible actions and change. Stakeholders must not only combine their efforts and reach consensus if sustainable change is to be implemented. They also hold vital information and insights into past experience of similar efforts, what went wrong and right and what could be done to bring about improvements in the future.

Stern describes the steps that might be followed by evaluators to involve stakeholders in a constructivist logic to lead to commonly perceived reasons for changes and solutions for reforms ⁽²⁵⁾:

- (a) 'identify the different stakeholders who potentially have a stake in these areas of concern;
- (b) conduct a series of initial discussions to clarify what they know, what they want and what are their interests;
- (c) feed back to all stakeholders their own and each other's interests, knowledge and concerns in a way that emphasises the similarities and differences;
- (d) clarify areas of agreement and disagreement and initiate discussions among the stakeholders and their representatives to clarify areas of consensus and continuing dissent;
- (e) agree what other sources of information could help move the stakeholders forward – perhaps by synthesising other available studies, perhaps by initiating new studies;
- (f) reach the best possible consensus about what should be done to improve VET provision and participation for the groups concerned.' (Stern, 2004, Section 4.3).

Therefore, stakeholders should agree first on strategies and areas for intervention. Second, during the change process, strategic phases to

meet, reflect and learn should be arranged for social actors and stakeholders (not only from the VET system itself but also from other systems). These are opportunities for giving feedback, correcting and improving internal and external consistency of a reform or policy.

Preskill and Torres (1999) in the context of organisational learning, underline that 'learning from evaluation will most likely occur when evaluation is collaborative [...] and builds communities of evaluation practice.' Therefore, communities of both evaluators and users should be developed if evaluation is to become part of an active learning process.

In the approach described, evaluation is seen as a tool for generating a continuous critical reflection process at various stages of reform. 'It emphasises the responsive, interactive, dialogic and "orchestrating" role of the evaluator because the sources of data that are privileged are seen to reside with stakeholders [...]' (Stern, 2004, Section 4.3). Evaluations are constructivist because they put the actors at the centre of evaluation rather than its outcomes (Viertel et al., 2004, Section 5.4), they are formative because they put emphasis on the learning process and dialogues necessary to change ⁽²⁶⁾.

2.4.2. Using building blocks to manage dialogue

For planning system reforms, the building blocks proposed and used by the European Training Foundation (ETF) (ibid., Section 2.2) are a functionalist analysis which can assist in defining in which part of a system to intervene. They constitute an attempt to provide a simple and transparent vehicle for managing the dialogue between key actors in a given country. In that respect, they are a tool to involve stakeholders in a constructivist logic, articulating their debates around elements or 'blocks' of the system.

Building blocks try to balance the complexity of VET systems and possible levels

⁽²⁴⁾ Philosophies and their use in evaluation are described in details in Part 1, Chapter 2.

⁽²⁵⁾ See also Part 1, Section 2.1.4 for more details on the constructivist philosophy.

⁽²⁶⁾ The evaluation accompanying the implementation of the Danish Reform 2000 is an example of how these principles can be put in action. The Danish case is described in details in Section 3.4.1.

of intervention with simplicity and transparency in the use of tools. They have been used by the ETF in transition countries as an operational model to analyse structures and practices, to make proposals for change in a consistent way, and to achieve common agenda among ministries and agencies. Building blocks, the way they operate and how they were applied by the ETF is explained in Box 3.3.

Box 3.3. The 'building blocks' approach

In the building block approach, eight topics offer an operational model for analysing existing structures and practices:

- educational management and administration;
- curriculum, assessment and certification;
- financing of VET;
- the labour market and social partnerships;
- educational standards and quality control;
- in-service teacher training;
- legislation;
- labour market and adult education.

Each topic is analysed in detail in a structured and systematic way by the actors involved. For example; the following three questions were used as a guide for the building block on legislation:

- should there be separate or integrated legislation for general education and VET?
- should existing legislation be left largely unchanged with modification of regulations only or should there be clear, new legislation, indeed an educational reform act?
- if new legislation is declared necessary, should it be short, simple and transparent, leaving the details to regulation?

By analysing different elements of a system, a common conceptual grasp of the issues at stake is attained. A common language is also established and used by a relatively large group of key actors to discuss structures, functions, and institutions of a VET system to be transformed. Encompassing all elements of the system, the building blocks approach was qualified by the ETF as a useful tool for designing well-grounded and specified VET reform strategies.

Source: Viertel et al. (2004, Section 2.2).

According to Viertel et al. (2004, Section 3.2), building blocks can also be used to assess how well a system performs. They can help in

formulating, for the various elements of the system, 'ideal practices' that can be compared to actual ones. This exercise would be influenced by the values and perceptions of what makes a 'good system' (see criteria-based evaluation in culturally bounded frameworks in Section 5.2). Therefore, while using the building blocks method in evaluation, cultural and other values of the evaluators and stakeholders need to be made explicit.

The evaluator has to be aware that the building blocks remain a functional analysis of parts of a VET system. Alone, this does not make one capable of fully understanding internal and external consistency, nor historical, cultural, structural and institutional particularities or the dynamics of each VET system.

2.4.3. Looking for explanations

Building blocks need to be complemented, first, by methods which would allow the evaluator to grasp the nature of the relationship between different parts and the totality (looking for internal and external inconsistencies, see Section 2.3). Second, structural-historical approaches, provide explanatory frameworks for the evaluator to grasp the historic roots and the dynamic relationships. These are necessary to understand the system logic and the possible driving forces for change and therefore possible strategic levers (Viertel et al., 2004, Section 3).

A structural-historical analysis starts with an overview of existing practices, functions and structures, as they appear to the evaluator. The historical origins of the main structures are then traced back to understand how a specific phenomenon was established (*ibid.*, Section 7). Potentially useful explanatory frameworks when looking at VET systems are the following:

- (a) historical and cultural background of different institutions and practices in contemporary VET systems, for example the profusion of independent VET providers in Denmark derives from the Grundtvig and the free-school and folk-high school tradition;

- (b) structuralist explanations, for example the specific links that existed between the economic, social and political systems in former communist countries are making the adaptation of VET to the demands of the market economy very difficult;
- (c) systemic explanations, to understand how changes in one part of the system might interact with and influence others, as explained in more detail in the discussion of internal and external consistency in Section 2.3.
- (d) economic, labour market or social 'laws', for example the changes in supply and demand (see Chapter 1), the dominant mode of regulation of the labour market (internal, external, occupational labour market; Descy and Tessaring, 2001a), social dynamics and mechanisms of social reproduction, etc.;
- (e) functionalist explanations, for example the functions socially attributed to education and training, and in particular VET, such as socialising and developing citizens, qualifying the labour force, selecting the elite, etc.

Understanding the system logic also requires a deeper insight into what are, have been or could have been, strategic levers for change, i.e. it requires an understanding of the organisational and institutional processes of change (Viertel et al., 2004, Section 5.1).

Designing appropriate policies and implementing them successfully means applying the building blocks approach, complemented by understanding the contexts of change (i.e. what will make the blocks hold together).

2.5. Assessing education and training systems and reform strategies ⁽²⁷⁾

In this section we shall look specifically at the kind of tools that are of interest in assessing the performance of education and training and VET systems and/or reform strategies. First, we make a few remarks on the objectives of summative evaluation of systemic VET reforms.

Although evaluation criteria such as effectiveness, efficiency, impact and sustainability, are relevant for project evaluation, they need to be re-examined and adapted if the subject of evaluation is not a single project but a system. Eventually, new criteria, tools and methods are to be explored.

Applying a summative approach to education and training and VET systems aims to identify where a system or a country stands at a certain point in time and to propose corrective measures or future possible developments. In this context, we are not discussing the results of an intervention that started at time x , stopped at $x+1$, concerned a very specific target group, under very specific conditions and which remained isolated from other elements. Systems are dynamic and policy formulation is a process.

Methods for assessing VET, studying and measuring the results of reforms face the challenge of deciding:

- (a) what to measure within the multiples outcomes, results and changes produced;
- (b) what objectives are to be considered within the numerous stated and non-stated objectives of a reform;
- (c) how to measure, at what level, with what kind of tools;
- (d) how to assess whether the changes, results, outcomes measured are a direct result of the policy.

Coles (2004, Section 7) recommends making the following considerations to improve the measurement of reform impact:

⁽²⁷⁾ The tools and methods described in this section are not only used in evaluating reforms. They can also be applied for programme evaluation. Nevertheless, Coles (2004) identified them as common practices while evaluating reforms in Europe. In this contribution to the background report, Coles reviews 19 evaluations of reforms, reports and draw some conclusions on practices in Europe. It is the main source for this section of the report.

- (a) identify what is driving the reform; this can clarify the need for impact analysis;
- (b) clarify the purpose of the evaluation programme: is the quest for effects on individuals, enterprises, wider communities adequately represented in the purpose of the evaluation?
- (c) ensure the objectives of the evaluation cover the desired outcomes of the reform: check that the data likely to be generated can be analysed to meet the objectives of the evaluation and are meaningful regarding the objectives of the reform;
- (d) consider methodology and use opportunities to make the process as rigorous as possible, i.e. considering baseline measurement and the timing of evaluation carefully;
- (e) if possible, analyse the effects of variables that cannot be controlled, for example changes in the labour market;
- (f) take all levels of impact into account, individuals, institutional and macro level.

Even when applying a summative approach, the evaluators have to look not only at measuring and comparing, but also at understanding whether the results of the evaluation are to serve the development either of corrective or future policies. This is illustrated through the different possible tools that are presented below.

2.5.1. User questionnaires and surveys

Coles (2004), in his review of 19 evaluation reports, identifies user response questionnaires as the most common evaluation practice in assessing reform impact. These can be addressed to participants, institutions or various other stakeholders (e.g., employers). The questionnaires are usually structured in three main parts (Coles, 2004: Section 4.2):

- (a) personal and demographic background information of respondents;
- (b) rating by respondents of aspects of the VET provision they are (or have been) participating in;
- (c) judgements of respondents about future

prospects for employment and learning progression.

The individual is the focus of many reforms and his/her views are sought in many evaluations (Coles, 2004: Section 6.3). A common practice is then to focus on satisfaction and other subjective measures. The individual is asked about his/her degree of satisfaction concerning different aspects of the intervention, sometimes in comparison with the situation before (e.g. improvement) ⁽²⁸⁾. The reasons for finding the intervention attractive, for dropping out and its relevance to employment are also common topics on which individuals' opinions can be useful for the evaluator. He/she can also be asked how he/she feels relative to others (not participating in the intervention), in terms of career chances for example. Participants and non-participants can also be asked about their opinion on different aspects of a specific intervention.

As opposed to views and opinions, individuals can also be asked about their (objective) situation, past and future, such as employment status, income, etc. ⁽²⁹⁾. This becomes most meaningful in cases where (changes in) characteristics of the intervention group are compared to a demographically matched group over a period of time (see Part 2, Section 1.3.1.2 for more details on control group methodologies).

Individuals are often the main targets and the main recipients of reforms. It is the benefits they gain (in terms of learning, skills, employment situation, well-being, etc.) and their behaviour (progression to higher education, attitude towards learning, entrepreneurship, job search, etc.) that are at stake while reforming. Therefore, they constitute a key subject for evaluation. Effects and outcomes at individual level can be determined by various measures ranging from the purely subjective to more objective/scientific approaches. All are valuable but they do not fulfil the same requirements in terms of using the summative evaluation results for

⁽²⁸⁾ This is to be related to the 'before-after' kind of method, for more details see Part 2, Section 1.3.3.3.

⁽²⁹⁾ This is to be related to the 'before-after' kind of method, for more details see Part 2, Section 1.3.3.3.

improving and correcting features of a policy⁽³⁰⁾.

Organisations and institutions are essential to reforming a system. If institutions and other organisations (enterprises, education and training providers, regional or local government, etc.) fail to absorb and respond to change, a new policy is due to fail. It is difficult, nevertheless, to measure effects and impacts at organisational and institutional level. This is the reason why this level, and the way it reacts to a reform, is often used as an explanatory variable for the successes and failures encountered at individual level. Nevertheless, developing knowledge on how institutions have reacted and adapted to a reform is also a key purpose of evaluation (Coles, 2004; quoting Stern, 2004).

Inquiring about the views of enterprises and employers on the impact of a reform does not inform on its outcomes at the level of the enterprise itself (in terms of productivity for example). In some cases, enterprises are involved in the part of the VET system under reform, in others they are simply the main users of VET results. In practice, enterprise consultation usually takes the form of satisfaction surveys. For example, their views can be surveyed on different aspects of a training course (including quality) or about the qualities and skills of the graduates of a part of the VET under reform compared to those participating in traditional pathways. This kind of survey, when very specific, can also help to identify issues where VET could be improved. The involvement of the employers themselves in a reform can be an indicator of its success and impacts, as well as their wish to continue to be involved in a VET initiative.

Coles (2004) does not mention any attempt to measure the effects of reforms, even in terms of satisfaction, at VET provider level. This is surprising given that they are at the heart of change. Aspects that are often neglected while evaluating the impact of a reform are effects on:

- (a) inter-institutional relations (e.g. between VET and general education institutions);
- (b) intra-institutional relations (e.g. between managerial and teaching staff);
- (c) teachers' roles, functions, careers, training, etc.

They can be considered as of non primary interest because they are not the primary target for change. In the Danish case presented in Section 3.4.1., the impact of the reform at institutional level, on the roles of managerial and teaching staff, has been integrated in the evaluation and even used as a lever for change for the new policy.

Questionnaires to the different stakeholders of a reform, therefore, appear to be useful tool in terms of gathering information on impact but also explanatory variables. These questionnaires can be targeted at the individual and the institutional level and might focus as well on subjective and objectives issues. The results and the information obtained by these questionnaires can be useful in formative and in summative perspectives. They allow for obtaining feedback from the main reform stakeholders.

2.5.2. Benchmarking

In education and training systems evaluation, and more generally in policy evaluation, benchmarking is an approach which is increasingly being applied. Viertel et al. (2004, Section 6.3) give a simple definition of benchmarking. It is the 'process of learning by making comparisons'. It is used as a formal and deliberate process of comparison aimed at detecting weaknesses and creating ideas for improvement. Benchmarking is a tool used in mutual policy learning at EU level and beyond.

When establishing benchmarks two methods can be followed:

- (a) norm-referenced: the benchmark is the best performer, with matching performance becoming the norm (e.g. the benchmarking exercise carried out in Denmark, Section 3.3 and Box 3.4).

⁽³⁰⁾ This variety of measures and tools shows that the demarcation between qualitative and quantitative methods is blurred. Qualitative finding from questionnaires and surveys are sometimes transposed into quantitative figures.

(b) criteria-based: the performance is fixed according to criteria which are chosen independently of any best performance (e.g. halving the number of early school leavers until 2010, compared to 2000 in the EU Member States, Section 3.3 and Box 3.6) ⁽³¹⁾ ⁽³²⁾.

Progress or the success is then judged according to the norm or the chosen criteria and a comparison is used to identify options and to help progression towards the best performance or achievement of the criteria. Therefore, benchmarking is organised around establishing indicators, the designation of best performers and best practices and the development of a strategy to reach the set benchmark.

For benchmarking education and training systems, the first main obstacles are the development of appropriate indicators and the availability of data. In addition, a careful reflection on the strategy to be developed to reach the set performance level is needed, as well as on the coherence between benchmarks and indicators adopted. A simple transfer of the strategies adopted by the best performers is questionable. Crouch et al. (1999, cited in Nieuwenhuis and Shapiro, 2004) do not believe in a single best solution at system level. Copying best practices from elsewhere needs careful planning and consideration of the institutional history of both source and receiving systems.

2.5.2.1. *Culturally-bounded evaluation frameworks*

There are no standard models for of a 'good' VET system. There are common challenges but different priorities, institutional structure and educational histories. When developing criteria-based evaluation methods, benchmarking and best practice exercises, evaluators have to be aware of the culturally-bound conceptions they include. These frameworks must be applied

consciously, in accordance with the actors of the system, when proceeding to evaluation. Examples of currently common culturally-bounded frameworks used when evaluating VET policies and systems are:

- (a) the lifelong learning concept, now recommended everywhere. System development tends increasingly to be evaluated in connection with this concept of reference, advocated by the main international institutions (European Commission, OECD, Unesco, World Bank, etc.). It is even so in countries where the rudimentary conditions for initial VET are not in place;
- (b) the modularisation trends: advocates of output-based, modularised and credit approaches tend to forget that this conception of VET systems comes from the Anglo-Saxon countries which, because of their dominant governance model for education and training and modes of regulation of the labour market, had no other choice than to develop systems with little emphasis on process (curricula, teachers, pedagogics, etc.);
- (c) the so-called 'modern' high-order competences which are promoted as basic skills for the 'knowledge' economy (adaptability, responsibility for one's own learning, etc.) undermine the category of skilled workers. They are also based on theoretical representations of the labour market which are, in many cases, a caricature of reality (Viertel et al., 2004, Section 3.3).

Although the applicability of benchmarking is controversial, and the risk of applying blind and mechanical 'print-out' of 'good' systems and culturally-bounded models is high, it is a method that is more and more widely used internationally and in the EU. In this context, it has the merit to open up political and national debate ⁽³³⁾.

⁽³¹⁾ However, any criterion is normative.

⁽³²⁾ Another possible approach is to use the elements of a theory as standards for comparison with actual practices. Struyven and Steurs (2004; see also Section 4.2.1.1.) for example use the theory of quasi-market developed by Le Grand and Bartlett (1993) as a frame of reference against which actual strategies of reform of employment services are assessed.

⁽³³⁾ In the framework of reaching the ambitious goals set for the EU by the Lisbon strategy, a number of benchmarks have been developed. They are discussed in more depth in Section 3.3.

2.5.3. Macro level

Although links and correlations can be established between education and training and economic and social indicators at macro level (i.e. in society as a whole) ⁽³⁴⁾, it is much more difficult to measure the effects and impacts of specific reforms or how well systems perform. This is regrettable because it is at macro level that many of the drivers for change can be found. At aggregate level though, some indicators can be built to evaluate, for example, the success of the transition between school and work, the employment status of graduates from different pathways, or even the rates of return of various pathways. These can be used as indicators of the success and impact of a reform ⁽³⁵⁾ but they are weak at identifying areas for improving education and training and VET systems.

Wolf (2002) questions the generally established relationship between macro level indicators and education and training systems, and especially the implications for education and training policy. She gives an example, using a very common indicator: the more educated tend to earn more. The natural policy conclusion from this observation would be that to in order to increase earning levels of the population we must increase individual educational levels. '[But] we cannot conclude that if everyone had the same education as top earners, they would have the same income; and the more we expand and lengthen further higher education, the less reason we have to claim this.' ⁽³⁶⁾ (Wolf, 2002: 14). The indicators built at macro level to measure the achievement of certain goals and targets are quantitative and are, therefore, poorly suited to indicating where to undertake qualitative system change. They have, however, the advantage of opening debates, as in the case of international surveys such as IALS or PISA.

Policy design and learning, however, requires much more refined methods if interventions and reforms have to be designed ⁽³⁷⁾.

2.5.4. Evaluation timing: short-term evaluation for long-term change

Is there a proper time, or elapsed time, to measure the effect of a reform?

It appears that there is no real answer to this question. First, it must obviously be adapted to the object of study. What might sound like a proper elapsed time for looking at the outcomes of one reform, might not be appropriate at all for another. Second, different points in time might be chosen to evaluate different kinds of effects. Third, long-term evaluations of effects are more appropriate to system change but they often do not fit the short-term policy-making cycle. Some long-term analysis of results is taking place in a few cases. Finally, if one measures the effect at different points in time, the changes in the context (and especially the situation in the labour market) should always be used as possible explanatory variables.

Typically (according to Coles, 2004: Chapter 4), data will be gathered, analysed and reported but, in the meantime, the reform has moved on (policy and practice are rarely based on hard facts...) and the data appears to be, at least in part, redundant.

Established systems have an impact in their current form, no matter how well they perform. Summative evaluation is well-served by comparing the situation after the implementation of a policy or strategy of reform with the situation before (for the group of people concerned, e.g.). This requires the establishment of a baseline against which progress can be compared. A baseline can be established *a posteriori* and data gathered accordingly. However, it is most helpful if a baseline is established *ex ante* and the

⁽³⁴⁾ Part 4 of this report deals in detail with this issue.

⁽³⁵⁾ For the evaluation of active labour-market policies, methods have been developed to measure the net effects of a policy on the economy, i.e. the balance between benefits and negative effects (e.g. substitution, replacement, etc.). For more details on this method, see Part 2, Section 1.3.4.

⁽³⁶⁾ If higher skills have a positive impact on growth, as confirmed by most research, this implies a rise in earnings for all. However, if skills become more egalitarian, new criteria will emerge as dividers between workers or as selection criteria for recruitment.

⁽³⁷⁾ Part 4 discusses in detail macro level indicators and studies aimed at detecting the impact of education and training systems. Part 2, Section 2.1.1 discusses methodologies to determine the impact on growth of education and training.

appropriate data gathered at the stage of planning ⁽³⁸⁾.

Establishing a baseline and gathering data *ex ante* requires the planning of the evaluation and its integration in the reform strategy process. In addition to providing a better information base for evaluation, this might be a support to a precise formulation of the objectives of the exercise and of the indicators that will be used to assess their attainment.

Evaluating change, i.e. to be in a position to look at a given situation before and after, is at the heart of evaluating the impact of system reform. In this context the establishment of a baseline and the gathering of data, not only after, but also at planning and implementation stage of a reform, are essential.

2.5.5. Interdisciplinarity: grasping the complexity

The constitution of interdisciplinary evaluation teams composed of researchers/evaluators from different fields of social research (economy, sociology, pedagogy, etc.) – and from different countries – is an asset when evaluating complex VET system. It can contribute to a holistic view of change.

This is the common practice adopted by the OECD in the framework of the policy reviews they organise (see Section 3.2 for a detailed description of the policy reviews process and especially of the adult learning thematic review). The method adopted here is a form of peer review, where the evaluation is carried out by a team of experts in education and training during a country visit. The OECD pays special attention to constituting teams of experts that are balanced in terms of competence, country of origin, and even of gender. The aim is to ensure good coverage in terms of approaches to the aspect of the system reviewed. It should also be underlined that the OECD reviews are organised to ensure the active involvement of all the stakeholders in the system under review. The practice of using interdisciplinary

and international experts is also common in evaluating Phare VET programmes and the Leonardo da Vinci programme (see Section 3.6).

The evaluation of Dutch VET system reform (presented in detail in Section 3.4.2) was an assessment of the outcomes of different reform goals and strategies by different teams of researchers. Each of them had a discipline background adapted to the object of research (links between VET and production system, response to social demand, quality of educational process, efficiency, autonomy and quality assurance at institutional level, etc.). While the Ministry of Education drew up the questions to steer the evaluation, the role of each team was to translate these into a set of measurable variables and to draw conclusions on the achievement of the reform's goals to assist political evaluation. A common evaluation report and recommendations were then drawn up on the basis of the analysis and recommendations made by each team.

The complexity of VET systems with its many actors (institutional or not), calls for multi-disciplinary and multi-cultural evaluation teams. Furthermore, as Viertel et al. underline (2004, Section 3.2), evaluators cannot extract themselves from their own history, cultures and scientific background. In multidisciplinary teams, room is created for interaction and learning from these difference.

2.6. Conclusions

This chapter discussed possible approaches to, and use of, evaluation of education and training systems and reforms:

- (a) how can evaluation be used as a steering device, to accompany change? How can it provide timely feedback on the way a reform, in interacting with the rest of the system(s) and actors, achieves its goals and complies with initial policy aims? How

⁽³⁸⁾ Another possibility is to compare the situation after a reform for the group participating/affected to the situation of another group having the same characteristics but not affected. However this method is more easily applied to programme evaluation. The very nature of a system reform is to be large scale (unless it is a pilot phase) and therefore finding a matched group not affected (not even indirectly) might be difficult. For more discussion on the various evaluation methods, see Part 2, Section 1.3.3.

can evaluation be used to correct and adjust internal and/or external inconsistencies?

- (b) how can evaluation support policy decision, establishment of priorities, choice of objectives, means and level(s) of action and planning of implementation?
- (c) how can evaluation be used to assess and judge the results of a reform, how can it help to assess where a system or a country stands? How can it feed into the policy cycle and contribute to policy learning?

Policy-makers are expanding their views about the nature of useful evaluation and are increasingly asking for information on system performance. A shift towards systemic approaches can also be observed in the evaluation practices applied to education and training and VET. However, while a fairly robust evaluation methodology exists for programme evaluation, this is not the case for system evaluation.

VET is a complex social system. Reforming it requires interactions, debates and compromises between institutions and actors of two major areas of life, learning and working. Reforming requires analysis of relationships, communication channels and roles of the various stakeholders. It should also rely on the understanding that changes in one component will lead to change in other components and in the system as a whole. A successful reform strategy takes into account internal consistency between the different elements of the system as well as external consistency, i.e. understanding and caring for the way changes will affect other systems.

In addition to assessing and judging the performance of an education and training system, or whether a reform has achieved the initial goals ascribed to it, evaluation can assist policy-makers and support the planning and implementation of change. Here, formative and constructivist evaluation design are virtually indispensable. In the context of education and training systems, evaluation is therefore rarely singularly formative or summative. Evaluation, rather than being distinguished according to the two approaches, should be seen a tool to be used continuously in the policy cycle and

that can fulfil different functions.

Constructivist thinking and the more formative and participative evaluation approaches are particularly useful in contexts where there are many stakeholders and the entire system will not be able to progress if there is no broad consensus. They should be applied when deciding and planning reforms strategies but also during implementation, to allow for mutual learning, feedback and improvements. Evaluation is then a tool for generating continuous critical reflection at various stages of a reform.

When fostering system change, building blocks can be used to manage dialogue with key actors from inside and outside the system. Building blocks can be compared to the architectural elements of a VET system. They are initially sufficient to engage in a dialogue, providing a frame of reference towards which the different elements of the system can be assessed. However, they need to be complemented by methods which reflect the systemic nature of the relationship between the parts and their totality, the historical, cultural, and institutional roots and the 'logic' of the system. In that way, evaluators might avoid blindly and mechanically applying hobby-horse explanations and so develop, in a community of VET practitioners, a new attitude towards argumentation using the different explanatory framework provided to them.

Applying summative evaluation to education and training and VET systems means trying to identify where a system or a country stands at a certain point in time and propose corrective measures and future possible developments. Evaluation criteria such as effectiveness, efficiency, impact and sustainability, relevant for project evaluation, are to be examined again if the subject of evaluation is a system. A proper evaluation methodology for systems has probably still to be elaborated. However, various tools can be used (user questionnaires and satisfaction surveys, benchmarks, macro level indicators) and recommendations can be made (forming multidisciplinary teams, establishing baselines, measuring long-term outcomes).

Large-scale *ex post* evaluations, carried out (years) after a new policy has been

implemented, are interesting for understanding the process of change and the policy results at different levels. As a learning tool, they are very useful for social researchers but they often come too late from a policy point of view.

Summative evaluations of system reforms, to reach their potential for policy learning and improvement, are most useful when used after a pilot phase (before a new policy is implemented full-scale), in collaboration with the main reform partners (institutions, teachers and trainers, individuals/participants). It is then an opportunity to identify success, failures and areas for improvement. In this case, it can feed into the reform and the policy formulation process.

The British white paper *Modernising government* identifies learning from experience as a key in policy-making: 'Government should regard policy as a continuous learning process not as a series of one-off initiatives. We will improve our use of evidence and research so that we understand better the problems we are trying to address. We will make use of pilot schemes to encourage innovations and test whether they work' (cited in Nieuwenhuis and Shapiro, 2004: Chapter 3).

The next chapter is dedicated to various evaluation practices and case studies that involve a system perspective. It includes concrete examples where the various tools and methods discussed in the present chapter have been applied.

3. Policy evaluation and policy learning in action ⁽³⁹⁾

3.1. Introduction

The principles and models to apply for evaluating education and training and VET systems presented in the previous chapter were partly drawn from the evaluation studies gathered for this research report:

- (a) the OECD policy review on adult learning;
- (b) benchmarking and learning from good practice in the EU;
- (c) Danish and the Dutch evaluations of national VET system reform;
- (d) the role of teachers, trainers and VET professionals in Croatian VET reform;
- (e) evaluation of Community programmes in education and training.

In this chapter, these examples, plus lessons and recommendations that can be drawn from them, are presented and discussed. We will conclude with general remarks on evaluation in supporting policy-making and policy learning.

3.2. Evaluation in an international context (the OECD policy reviews) ⁽⁴⁰⁾

The OECD has carried out various thematic national policy reviews in education. They have touched upon varied subjects such as transition from school to work (OECD, 2000), pre-school education (ibid., 2001c), and, more recently, adult learning (ibid., 2003a). These reviews rely on voluntary participation by 'client' countries (who finance their own review) ⁽⁴¹⁾, criteria set by the OECD Secretariat

⁽³⁹⁾ Chapter 3 is based, amongst other sources, on the following contribution to the Background report: Pont and Werquin (2004); Viertel et al. (2004); Nieuwenhuis and Shapiro (2004); Baumgartl et al. (2004); Lauterbach et al. (2004); Molsosa (2004).

⁽⁴⁰⁾ This section is based on the description of the policy review on adult learning and its impact on national policies carried out by the OECD (provided as background report by Pont and Werquin, 2004).

⁽⁴¹⁾ Each OECD country is free to decide whether or not it wishes to take part in the work. Evaluation of the reforms, therefore, has to take into account a mixture of causes and effects:

- (a) either the country has already launched reforms and the thematic review merely observes and approves them or not. There, the impact of the thematic review is probably marginal, but it may help to evaluate or validate reforms under way;
- (b) the country is in a period of preparation and reflection and then the thematic review recommendations are likely to have more of an impact;
- (c) the two are mixed (most often the case): the task of the thematic review experts is both to judge any reforms under way and to suggest what might work better. Pinpointing reasons for success or failure or assessing the validity of the recommendations then becomes problematic as a whole range of causes and effects are mixed.

and careful choice of experts evaluating countries' policies.

Thematic policy reviews can be seen as a form of a peer review in international comparisons. The main objectives are to examine under what conditions a policy may provide successful results to offer lessons for the future and for the countries being compared. In other words, they try to identify good practices and to measure their transferability.

A complex, multi-stage, process is being developed for this purpose (Table 3.2), with each stage designed to highlight what has made an experiment bear fruit. The purpose of this section is to describe this process as it was applied in the thematic review on adult learning (OECD, 2003a).

A thematic review relies on combining of a kind of self-evaluation carried out by the country under review (in the form of a background report) and external evaluation (carried out during a country visit) from a team of experts.

The first exercise a country has to undertake when participating in a thematic review is to provide a background report to the OECD secretariat and to the experts who will undertake the visit. In the case described here, this report had to describe the state of adult learning and to provide material on selected issues relevant for the thematic review. The background report must be descriptive. The list of issues to be covered is defined at OECD secretariat level. In the case of adult learning, these issues were:

- (a) context and stage reached in the thinking about adult learning;
- (b) access to adult learning;
- (c) learning motivations;
- (d) measures to bring training supply and demand into step;
- (e) adults' needs and the way in which they are being satisfied;
- (f) public action at all levels, whether national, local or municipal;

- (g) evaluation, whether economic or social, individual or collective;
- (h) recent initiatives and their respective impact.

The value of the background report lies first in that it offers the possibility of having two distinct sources of information on adult learning: the report itself and the synthesis note written after the country's visit. It, therefore, allows identification of eventual contradictions between the internal vision of an issue and the way it appears to external evaluators. Second, the background report allows extraction of explanatory variables for the observations that will be made during the visit ⁽⁴²⁾. It is a source for detecting, for example, historical, cultural and systemic factors that drive a system. It is, finally, an opportunity to gather quantitative data.

It is also important to mention that, in the case of adult learning, this background report is one of the few comprehensive documents available.

Following the background report, a country visit takes place. The visiting team represents, as far as possible, different fields of competences (education sciences, psychology, sociology, economics, etc.), different countries and is gender balanced. Relying on this 'multi' team is a way to ensure the representation of different backgrounds for carrying out the evaluation exercise. The visit lasts 5 to 10 working days and includes between 50 and 100 meetings and/or field visits. It leads to the writing of a synthesis note.

The four themes leading the adult learning review have been chosen by OECD Secretariat and are identical for all the countries involved:

- (a) how can the government, social partners and other actors involved improve the incentives and motivations of adults to learn?
- (b) how can the fragmentation of the adult learning system be reduced and replaced by an integrated approach?
- (c) how can the quality of training and the variety of supply be improved?

⁽⁴²⁾ Viertel et al. (2004), refer also to the need for identifying explanatory framework to allow the evaluator(s) to understand and explain the dynamics of a system, the role of the different actors, including the institutional ones and the possible levels of change (see also Section 2.4.3.).

Table 3.2. Stages of a thematic policy review

Stage	Characteristics	Added-value
Background report	Prepared in the country visited. Issues to be covered defined by the OECD. Describe the state of the art. Provide descriptive information and background statistics.	First source of information on the subject. Source for detecting e.g. historical and cultural factors at stake. Gather quantitative data. A form of self-evaluation is carried out at country level.
Pre-visit	Short (a day). From a member of the OECD Secretariat. To prepare the programme of the visit.	Clarify the objective of the exercise and ensure adhesion to the evaluation method. Ensure that all actors and stakeholders are involved in the country visit (all ministries, employers, unions, ONG, etc.).
Country visit	From a multidisciplinary and multinational team. 5 to 10 working days (50 to 100 meetings and/or field visits) ^(a) .	Measure the stage reached at field level. Discover inconsistency between a central discourse and local thinking and practices. Opportunity for building explanatory framework. Conducive to exchange and dialogue between actors and between actors and evaluators.
Preparation of a country synthesis note	By one member of the external expert team. Organised according to themes selected by the OECD Secretariat, common to all countries visited.	Tailored evaluation of national policies, taking into account the previous stages (background report and visit), from an external expert team.
Preparation of the comparative report	Prepared by the OECD Secretariat. Organised according to themes selected by the OECD Secretariat.	Comparative analysis of a wide range of information (country's self-assessment, visits, analysis of the national situation by external expert). Puts the country in perspective. Discusses good (and bad) practices and provide explanatory framework. Explain differences between systems or between apparently similar options.

^(a) In the case of the thematic review on adult learning
Source: Pont and Werquin, 2004, adapted by the authors.

(d) how can the cohesion and the efficiency of adult training policies be improved?

Country visits are crucial to the process because they allow the team to point out successful actions, to measure the stage reached in each field and, in particular, to find out whether there is a consistency between what is being said centrally and what is happening or being perceived locally. In an international perspective, a field visit is an opportunity for explaining differences between various systems or between apparently similar options and choices. They are designed to be conducive to exchanges and dialogues between actors in the country concerned and

between actors and evaluators. The main danger of bias is that it is organised by the country itself. It is, therefore, sometimes difficult to judge how representative are the people met or the practices presented. It also creates a bias towards good practice, which is regrettable because 'bad' practices are probably as conducive to learn from.

The synthesis country reports, very useful in themselves for each country concerned, lead to the preparation of a comparative report. This is prepared by the OECD secretariat. It puts the countries in perspective, showing different policies, explaining in what respect they function well or less well and under what

conditions. (Annex 4 presents selected results and recommendations from the review on adult learning extracted from that synthesis report; OECD, 2003a).

All the steps followed during the thematic review are designed to optimise the depth and the quality of the process, which is enriched by constant exchanges, during the visit and after, between the people in charge of the project in each country and the experts. Countries involved in the thematic review on adult learning have generally noted that this work has been very useful (some of the participating countries have mentioned their willingness to participate in a focused review ⁽⁴³⁾).

A thematic review on issues such as adult learning makes them visible (even to non-participating countries). It helps to place them at the top of the political agenda and is an opportunity for the OECD Secretariat to transfer and reinterpret the debate from the international level to the national one ⁽⁴⁴⁾.

The strength of a thematic review is that it is an independent evaluation led by external experts which provides reference points for national policies in an international context. It also obliges institutions, ministries, social partners and providers to combine their efforts to offer the most appropriate overview of the system at all stages of activities. Therefore, it becomes an opportunity for all actors to come together, to be participative in their own evaluation and to learn about their own system and initiatives. It generates collaboration between the various actors and mutual learning (even sometimes between policy-makers and practitioners).

For comparative research, and to the community of policy-makers, thematic reviews also offer data on policies and practices, presenting a wide panorama of national approaches to improving education and training.

A somewhat weak point of the methodology applied is that either the profile of the team of experts or the nature of the sites visited can influence the attention given to an issue. But this is intrinsic to many (if not all) evaluation exercises. Another weak point is that the framework of analysis imposed by the OECD Secretariat (see the four themes mentioned above) interferes with the evaluation and limits the degree of freedom with which the experts can look at and assess systems. Also it might not always be the most appropriate framework for some of countries involved, depending on the stage of development of their systems or policies ⁽⁴⁵⁾. It probably also influences the nature of the recommendations made ⁽⁴⁶⁾.

Pont and Werquin (2004), emphasise that this form of external evaluation can by no means go into the depth, details and contextualisation of an internal evaluation carried out at national level. For a complete evaluation of a system, the former can only complement the latter. It is noteworthy that one of the recommendations of the thematic review in all countries is the necessity to conduct more internal evaluation work.

⁽⁴³⁾ The first wave of the thematic review on adult learning took place between December 1999 and the end of 2002. It involved Canada, Denmark, Finland, Norway, Portugal, Spain, Sweden, Switzerland and the UK. The comparative report was published in spring 2003 (OECD, 2003a). During 2003 and 2004, a second wave is taking place. It involves the following countries: Austria, Germany, Hungary, Mexico, the Netherlands, Korea, Poland, the UK and the US (the UK was part already of the first wave and Germany, the Netherlands and the US have already expressed their interest in a focused review). The second wave offers an opportunity to verify and refine analysis made during the first and to assess in different contexts some hypothesis made. The aim of a focused review is to look at what happened between the first and the second visit and analyse better issues left (partly) aside during the first exercise (e.g. new regions, new groups of population).

⁽⁴⁴⁾ The emphasis given by the OECD to adult learning and lifelong learning is not an isolated effort. Many other international bodies such as the European Commission, Unesco, the World bank, ILO, etc., as well as many countries are using lifelong learning as the new guiding framework for education and training, and sometimes employment and social policies. This contributes towards creating a context where the potential impact of a thematic review and of a comparative analysis on adult learning can generate great interest.

⁽⁴⁵⁾ Countries, however, when engaging in the exercise, agree with this guiding framework.

⁽⁴⁶⁾ For recommendations to individual countries, the reader should either refer to Section 2.2.1, to the Annex 4, to the article of Pont and Werquin (2004) in the background reports or to the comparative report of the OECD (2003a).

Box 3.4. Applying benchmarking in Denmark

Denmark ranked as follows in comparison with France, Germany, Great Britain, Japan, the Netherlands, Sweden and the US in the benchmarking exercise carried out in 1999.

Above average

- number of hours of teaching per teacher in basic vocational education;
- pupil/teacher ratio in primary and secondary education schemes;
- small proportion of young people who have been unemployed for one year or more.

Average

- total expenditure per pupil in the educational system;
- annual number of hours of teaching of 12 year olds;
- proportion of population with upper secondary education;
- proportion of population holding tertiary education qualification;
- maths skills;
- natural science skills, upper secondary education;
- reading skills of 14 year olds;
- relative unemployment for 20 to 24 year olds;

Below average

- number of hours' teaching per teacher in primary and secondary education;
- annual number of hours' teaching of 14 year olds;
- natural science skills after eight years' schooling;
- reading skills of nine year olds.

Source: Viertel et al., 2004, Section 6.3, based on Finansministeriet, May 1999.

3.3. Learning by comparison in the EU

As described in Section 2.5.2, benchmarking is the process of learning by making comparisons. Specific performance indicators are developed and 'best performers' or criteria identified. Comparison is used to identify options to help progress towards the set goal.

Viertel et al. (2004, Section 6.3) explain how, in 1999, the Danish Ministry of Finance

used benchmarking to assess where Denmark stood as far as issues related to prosperity and social welfare were concerned.

The Danish Ministry chose a 'best performer' approach. Countries considered among the world's leaders in one or more relevant fields (France, Germany, Great Britain, Japan, the Netherlands, Sweden and the US) were selected. A number of indicators were then used to make comparisons between them for major areas of prosperity and social welfare. Box 3.4 presents the results of the comparison exercise for the education sectors, i.e. where Denmark stood compared to the two best or two worst of the group of countries. However the analysis of factors and practices leading to better results in best performing countries was not carried out. Furthermore, the measures chosen are only proxy indicators of the quality of education systems and of the preparation they provide for successful private and professional life.

Benchmarking has been useful in Denmark as a political exercise, opening national debates and identifying areas for improvement. Benchmarking methods, through international comparison on selected indicators, can therefore trigger national debates. The impact of results of the IALS or the PISA⁽⁴⁷⁾ studies on debates on education and training issues in some participating countries is also illustrative of the role indicators can play.

Within the EU, benchmarking is used as a tool to reach the goals set at the Lisbon Summit in 2000. It has been used in employment policy⁽⁴⁸⁾ and it is progressively being applied to education and training (see EC, 2001a; EC, 2002b). In this respect, the Commission and the Council have adopted a common work programme and defined the method of open coordination. Box 3.5 explains this method and Box 3.6 presents EU benchmarks in education and training.

⁽⁴⁷⁾ IALS: International adult literacy survey (OECD and Statistics Canada, 1997, 2000); PISA: Programme for international student assessment (OECD, 2001a)

⁽⁴⁸⁾ The use of benchmarks in social policy at EU level was initiated in the context of the process started in Luxembourg in 1997 and called the European Employment Strategy. Employment benchmarks are now an integrated part of the Lisbon Strategy (Extracts from Presidency conclusions on the Lisbon strategy by themes. European Councils: Lisbon to Brussels, March 2003).

Box 3.5. The open method of coordination

While respecting the breakdown of responsibilities envisaged in various EU treaties, the open method of coordination is a way of spreading knowledge of best practices and achieving a greater convergence towards the main EU goals. It is a new form of cooperation for the Member States based on a fully decentralised approach using variable forms of partnerships and designed to help them to progressively develop their own policies. It is based essentially on:

- identifying and defining jointly the objectives to be reached (the benchmarks, see Box 3.6);
- commonly-defined yardsticks (statistics, indicators) enabling Member States to know where they stand and to assess progress towards the set objectives;
- comparative cooperation tools to stimulate exchange and dissemination of good practices.

The exchange of experiences and the identification and selection of good practices are carried out in three strategic areas, corresponding to the 13 'concrete future objectives of education and training systems' (EC, 2001a; EC, 2003d). Eight working groups, supported by the Commission (DG EAC), are operating (the so-called 'objective groups'), covering all 13 objectives. They are composed of national experts appointed by the participating countries (Member States, EFTA and candidate countries) and representative of education and training stakeholders.

Source: http://europa.eu.int/comm/education/policies/pol/policy_en.html#methode; detailed work programme on the follow-up of the objectives of education and training systems in Europe (Council of the European Union, 2002) - EC, 2003b.

In the EU, the main obstacles for benchmarking in education and training are the development of appropriate indicators and the availability and comparability of data. Furthermore, various concerns may be expressed about the approach taken:

- using the EU average as a benchmark implies that if the best performing countries make further progress, this will 'relieve' others, less advanced from the 'burden' of investing in education and training over-proportionally;
- the set target refers to a non-weighted EU average, i.e. it does not take into account the size of the population in each country.

Box 3.6. EU benchmarks for education and training

To achieve the Lisbon European Council's strategic goal to make EU by 2010 'the most competitive and dynamic knowledge-based economy in the world capable of sustainable economic growth with more and better jobs and greater social cohesion', the heads of States and Governments agreed on some common objectives for education and training in the EU, within the overarching framework of lifelong learning:

- improving the quality and effectiveness of education and training systems in the EU;
- facilitating access of all to education and training systems;
- opening up education and training systems to the wider world.

Subsequently, the 2002 Barcelona European Council set a new goal 'to make Europe's education and training systems a world quality reference' by 2010.

The work programme for achieving these objectives foresees that progress will be monitored against agreed indicators expressed as average level performance of the EU15 Member States and of the three best performing ones.

Consequently, the Education Council agreed in May 2003 on the following European benchmarks:

- 'By 2010, all Member States should at least reduce the rate of early school leavers, with reference to the rate recorded in the year 2000, sufficiently to achieve an EU-average of 10 % or less.
- By 2010, the total number of graduates in mathematics science and technology in the EU should increase by at least 15 % while at the same time the level of gender imbalance should decrease.
- By 2010, at least 85 % of 22 year olds in the EU should have completed upper secondary education.
- By 2010, the percentage of 15 year olds with low achievement in reading literacy in the EU should have decreased by at least 20 % compared to the year 2000.
- By 2010, the EU-average level of participation in lifelong learning should be at least 12.5 % of the adult working age population (25-64 age group).'

In addition, 'the Commission invites Member States to continue to contribute to the achievement of the Lisbon objectives of substantial annual increases in per capita investments in human resources, and, in this respect, to set transparent benchmarks [...].'

Source: EC, 2002b; Press release 5/05/2003, available from Internet: <http://europa.eu.int/rapid/pressReleasesAction.do?reference=IP/03/620&format=HTML&aged=0&language=EN&guiLanguage=en> [cited 20.8.2004].

- Therefore, a progress in a large country affects the average more than a progress in a small country;
- (c) there could be some saturation level at which progress cannot be made any longer. In all societies, there will be always a group of people who are not able or not willing to study further (Tessaring, 2003);
 - (d) the various set benchmarks are treated as independent when they are, in fact, inter-related. For example the reduction of early school drop-out and the increase of the number of young people who leave the education system with at least the upper secondary level are two complementary objectives;
 - (e) countries' relative necessary progress is not considered. The rationale behind benchmarks is 'low achievers have to catch up' and in some cases there is a long way to go. Furthermore, is there really no progress to be made in best performing countries?
 - (f) another assumption hidden behind these benchmarks is the idea that 'since some education makes some of us rich, more would make more of us richer – or 'if two aspirins are good' it means that 'five aspirins are better'.' (Wolf, 2002; p. 28). This assumption is not necessarily true, substitution mechanisms and over-qualification, for example, tend to prove it;
 - (g) finally, the efforts to reach each education and training targets in each country should be converted into financial figures and additional support to countries where substantial financial investment would be required, should be considered (Tessaring, 2003).

In addition, when it comes to learning from benchmarking, the transferability of strategies adopted by the best performers is questionable. Crouch et al. (1999, cited in Nieuwenhuis and Shapiro, 2004) do not believe in a single best solution. Copying practices needs careful planning and consideration of both source and receiving systems.

Whether benchmarking can be applied to social systems is not at all obvious. In the context of employment policies, Schütz et al. (1998, cited by Viertel et al., 2004) think it easier to benchmark organisational entities (such as employment offices) than policies, whose success depends in part on other policies and institutions.

Therefore, although country benchmarking is rarely more than just an international comparison exercise and cannot be used for one country to learn directly from another, it has the merit of opening up political and national debate. In the EU, it provides common targets and frameworks of reference that need to be complemented by a considered and tailor-made policy for each context. Perhaps the solution is to develop tailor-made benchmarking, i.e. to define a common framework for the Member States but then to reflect on tailor-made targets, areas for improvement and desirable results for each country given its 'starting point'.

3.4. Evaluation of reforms of national VET systems ⁽⁴⁹⁾

In this section, evaluations of the reforms of national VET systems from Denmark and from the Netherlands will be presented and discussed. Both the reforms themselves and their evaluations are quite different. Reflecting on these will lead us to drawn some conclusions on the use of evaluation for supporting and assessing reform.

3.4.1. The Danish case: a 'reflective gathering of experience'

The reform (VET Reform 2000) of the Danish VET system was initiated, designed and implemented in ways that consciously aimed at fostering a learning process between policy formulation and implementation, using interactive/formative evaluation.

In Denmark, it was felt that a new legislative framework was needed for the VET system to

⁽⁴⁹⁾ This section is based on the detailed description of the Danish VET Reform 2000 and of the Dutch WEB reform and their evaluation by Nieuwenhuis and Shapiro (2004).

Table 3.3. Levels and elements of the Danish legislative framework for the Reform 2000

General axis		Individual axis
Central level	<ul style="list-style-type: none"> • legislation • main VET regulations • other regulations: core subjects, guidance, special assistance, etc. • educational regulations and administrative instructions • quality and assessment framework 	
Local level	<ul style="list-style-type: none"> • local educational plan • local conditions for the implementation of quality control and assessment 	
Implementation level	<ul style="list-style-type: none"> • the learning offer (modules) • regional school (supply) • teachers' teams planning • ICT infrastructure supporting modularisation and individualisation 	

Source: Nieuwenhuis and Shapiro, 2004, adapted by the authors.

cope with the changes in supply and demand for skills in the labour market (Chapter 1). The rationale was that continuous accommodation of the curricular structure and supply to new skill needs should be possible without necessitating constantly changing the legislative base. The reform also foresaw that the learning environment should allow learners to pursue their own goals, fostering 'learning to learn' competences (while traditionally skill formation in VET happens via instruction and imitation).

'Another priority is to improve the responsiveness of the system to the new qualification needs of the labour market. The labour market requires skilled workers at the right level of competence and with the right sort of training. The focus of the late 1990s is to grant more freedom to the individual learner. The VET reform provides for a system that offers students of all different types specific, individualised training. Thus, the main objective of the reform is to strike a new balance between the two aspects: employability by providing higher general personal and vocational skills required by the

business community and by establishing more flexible pathways to learning for the individual student.' (web site of the Danish Ministry of education) ⁽⁵⁰⁾.

A task force was set up prior to the reform. It was to translate elements from international research and experience into a legislative practice and to examine critically the existing barriers in the legislative framework. The task force's recommendations later set the stage for the follow-up and evaluation of the first pilot reform.

The main axis of the reform was to emphasise and legitimise individual students' active roles in the learning process by allowing them to shape their individual learning rather than to follow externally predefined pathways. This was made concrete by:

- (a) a modular structure of the curriculum;
- (b) local educational plans, allowing for local adaptation;
- (c) personal educational plans, documenting students' individualised learning pathways;
- (d) logbooks, documenting the students' workplace learning, meant as a link between school and workplace;

⁽⁵⁰⁾ Available from Internet: <http://pub.uvm.dk/2000/newstructure/4.htm> [cited 7.5.2004].

- (e) contact teachers, providing individual guidance to students and ensuring the links between individualised and collective pathways (the levels and elements of the Danish Reform 2000 are presented in Table 3.3).

To meet the requirements for a more flexible qualification structure, the various trades and professions were regrouped into six broad occupational categories: technology and communication; building and construction; crafts and techniques; 'from earth to table'; mechanics, transportation and logistics; and services.

Parallel to the reform, the Ministry of Education undertook a number of actions to create an environment receptive to change:

- (a) a study on new didactical paradigms in line with the intentions of the legislation (Shapiro and Christensen, 1999);
- (b) training of trainers' initiatives, in collaboration with the Danish vocational educational teachers' organisation;
- (c) school development projects on tools and practices that could support the change process (with the support of the National institute for vocational teacher training);
- (d) development of an ICT tool (web-based), in cooperation with practitioners and other actors of the reform, to support the school in managing the modularisation and the individualisation of the curriculum;
- (e) a gradual pilot reform (as from summer 1999) in a number of schools for better understanding of structural, organisational and pedagogical implications which might require further policy actions; and finally,
- (f) two evaluations of the pilot phase ⁽⁵¹⁾.

Before discussing these evaluations in more detail, we should stress that these actions probably played a role in increasing internal consistency while implementing the reform (Section 2.3). The various initiatives involved many actors, making them partners in the change process (especially teachers and trainers, Section 2.4.1).

3.4.1.1. *First evaluation (November 1999 – February 2000): main results and benefits*

This first evaluation was called by the Ministry of Education, a 'reflective gathering of experiences'. The aim was to inspire and assist the ministry, directors of schools, teachers, the National Institute for Vocational Education and local councils in further implementation of the reform. Its function was to identify barriers and difficulties as well as promising new practices, and to define where further policy interventions were needed.

Themes to be covered were:

- (a) the change perceived at work and in professional profiles and how it was dealt with;
- (b) knowledge of the reform and how the implementation was perceived;
- (c) the information and collaboration between the different actors;
- (d) the implementation models, qualification demands, frameworks and conditions;
- (e) the experiences with roles, instruments and structures;
- (f) the contact teacher: organisation and qualification for the new role.

The two main instruments used during the evaluation were:

- (a) a survey amongst school directors involved in the pilot and samples of teacher and student populations;
- (b) a complementary case study, using focus group interviews of coordinating fora at national, regional and implementation levels (teachers' unions, local and school councils).

The evaluation also included interviews with students.

The instruments were discussed with representatives of all actors to ensure the relevance of the process and commitment. This participation in the evaluation design led to a great deal of openness and readiness to share and debate the implementation process.

⁽⁵¹⁾ Many of these projects and publications compiled in the framework of the Reform 2000 have been web published by the Ministry of Education: <http://pub.uvm.dk>. *Forsøg med erhvervsuddannelsesreformen 2000: opsamling af de første erfaringer* (Pilot projects related to the VET Reform 2000: collection of first experiences) <http://pub.uvm.dk/2000/opsamling/>.

Beside the exploitation of specific results, the evaluation generated a climate and a dialogue between the participants, which helped in shaping further implementation efforts. For example, it appeared that the evaluation was in many cases an opportunity for a first meeting of a coordinating forum.

One interesting result of the evaluation was the comparison of centralised and decentralised approaches adopted at regional level. The centralised approach involved a consensus between the schools' directors on a common strategy at regional level while, in the decentralised one, learning by doing in each school was the preferred strategy. The former resulted in speeding up the implementation and developing a strong collaborative network amongst school directors but teachers felt insufficiently involved in the decision-making and were demotivated. The decentralised approach, in contrast, led to more motivation on the side of the teachers although it generated considerable delays in the development of pedagogical tools. As a consequence of these observations, in the second pilot round, the function of the directors as catalyst of change was stressed while paying attention to developing their role as carriers of debates and information. The National institute for vocational education even designed training programmes for middle management so that they could act as reform consultants.

This influence of the reform on the role of directors was reflected in the survey that identified a reshaping of professional identities of both middle managers and of teachers. The change of professional identity appeared to be for teachers a source of considerable stress. After the first three months, 25 % of the teachers surveyed considered looking for another job. The fact that they were becoming 'facilitators', rather than teachers in the

traditional sense, and that they had to work within a broader branch and technological perspective, were the source of this dissatisfaction. In the second pilot round, this was addressed by creating teachers' teams.

Some difficulties linked to individualisation were also detected. The question of how to continue accommodating the needs of different student groupings while also providing a reference group was raised.

Both surveys and interviews showed a high degree of satisfaction among students, of which the older ones especially thrived in the new environment with a high degree of autonomy. The contact teacher function appeared to be appreciated both by the students and by the teachers because it enabled more varied support strategies.

The results of this first evaluation were published (including on the web) ⁽⁵²⁾ and a one-day seminar and regional workshops were organised to debate the findings.

3.4.1.2. *Second evaluation (October 2000 – February 2001) – main results and benefits*

After the pilot, the reform was expanded (in January and again in August 2000) not only in terms of number of schools but also to all six occupational categories (the pilot covered only the basic parts of the programme). The second evaluation started in October 2000 and its aims were, as with the first exercise, to identify barriers and problems as well as successful practical solutions and strategies. Survey themes were built on the first exercise ⁽⁵³⁾ to be able to capture changes in perceptions and practices compared to the first evaluation. A sample survey of students that had dropped out from the programme/school was also constituted. This time the interviews focused on managers and teachers, apprenticeship companies (both apprentices and students),

⁽⁵²⁾ *Erfaringer fra forsøg med eud-reformen: grundforløb og hovedforløb 2000 (Experiences from pilot projects under the VET Reform: foundation and main modules)*. Available from Internet: <http://pub.uvm.dk/2001/eudreform2/> [cited 20.8.2004]. *Hvordan gik det?: sammenfatning af Erfaringer fra forsøg med eud-reformen i 2000 (How did it go?: summary of experiences from pilot projects in the VET Reform 2000)*. Available from Internet: <http://pub.uvm.dk/2001/eudreform1/index.html> [cited 20.8.2004].

⁽⁵³⁾ Themes covered were: organisational structures and changes; the pedagogical space; instruments and roles; qualification measures and organisational development; experiences with roles, instruments and structures; information and collaboration; companies and the apprenticeship scheme.

local educational councils and teachers' union representatives.

This second evaluation revealed much more flexibility in the duration of modules and in content and learning styles than was the case in the pilot phase. However, there were still unresolved problems related to clustering some of the trades into the six occupational categories.

For the students, the increased individualisation of pathways, although it appeared positive in many cases, created some difficulties for:

- (a) students choosing a prolonged basic programme because apprentice companies still had a notion of fixed duration programmes;
- (b) weaker students who were neither ready nor motivated to engage in jointly planning their educational routes;
- (c) some students thinking this flexibility had turned into a *laissez-faire* attitude from teachers not sufficiently challenging their choices.

After the evaluation, many schools started to focus on the development of the 'learning to learn' competence, i.e. students' ability to plan and select their own learning pathways. Because of the complexity of selection and planning of pathways, some schools decided to offer more and shorter options in the form of flexible duration projects.

New forms of assessment, linked to project-pedagogy seemed to be difficult to implement. After the second evaluation, the Ministry of Education initiated development work, in collaboration with the schools, on alternative forms of assessment.

Teachers, who were traditionally experts in a certain trade, had to become learning facilitators and to work together with other colleagues. This profoundly challenged their professional identity. Furthermore, the uncertainty as to whether the educational goals could be reached through the new learning style increased teachers' insecurity and threatened what they perceived as their role in 'safeguarding' the students. In contrast, the role of the contact teacher was appreciated and perceived as a success from

the point of view of teachers, students and managers. Finally, while teachers had gained more control over administrative and pedagogical planning from the reform, the financial functions tended to remain with the management.

The logbook and the personal educational plan were not used in a satisfactory manner. The logbook in particular was often perceived as a 'dead' file by the students, updated only at the initiative of the teacher/contact teacher. However, when used in a joint effort, it appeared to have enhanced the relationship between the apprentice, the apprentice company and the contact teacher. It allowed students to convey to the company what they had worked on in the previous period, making it easier for the company to plan the apprenticeship period. Students could also transfer some of their work experience when back at school. In that case, the logbook ensured continuity of interaction between the different actors. It also promoted good cooperation between schools and companies and a better understanding by the latter of the changes caused by the reform.

The Danish experience clearly shows how evaluation can be used to accompany reform, involving different actors not only as target groups for the evaluation but also in the evaluation design and, in doing so, developing a mechanism for (early) feedback and adjustments. The strength of the evaluation of the Danish reform was not only the material produced and published but the willingness and ability to present and debate findings and policy recommendations with different stakeholders. Used in that way, evaluation not only supports the achievement of the initial aims of the reform, it also permits learning from the actors involved at the different levels, from policy makers to students or teachers. This kind of evaluation is oriented towards accompanying, rather than providing 'proofs' about the effectiveness – or non-effectiveness – of new policies. It is a formative form of evaluation that seeks to influence and promote change. One criticism that can be made of this particular evaluation scheme, however, is that it emphasised identifying and correcting

internal inconsistency, while partly neglecting the external consistency elements ⁽⁵⁴⁾. It is, therefore, difficult to judge how the VET system changes can echo related systems, such as production.

3.4.2. Summative evaluation and VET system reform: the Dutch case ⁽⁵⁵⁾

A new law on vocational and adult education was launched in the Netherlands in 1996: the WEB ⁽⁵⁶⁾. It was meant to increase the efficiency and responsiveness of the VET system. The reform targeted mostly institutional and organisational change, also fostering regional and sectoral dimensions. Regional training centres (ROCs) having a large degree of autonomy were created and sectoral qualification structures were implemented.

The main features of the WEB are:

- (a) the unification of instruments and institutions: both initial VET and continuing adult education and training are provided by the ROCs;
- (b) an output-oriented qualification structure, caring for social and economic demands by providing basic qualifications for all and allowing the same output to be reached following different paths, either company- or school-based;
- (c) a regional perspective: the ROCs are supposed to cover all occupational training needs in their individual region through a merger of former sectoral vocational schools, apprenticeship support structures and adult education centres. They cover all sectors (except agriculture and some smaller sectors) and organise vocational learning with local companies;
- (d) a sectoral perspective: social partners, organised in 21 sectoral bodies, define the qualifications, which are then legitimised by the Ministry of Education. The ROCs apply for the delivery of certain

qualifications for their region and examination boards, connected to sectoral bodies, verify the quality of examinations and assessments;

- (e) ROCs autonomy: they can choose content, didactics and the organisation of practical learning and are responsible for the quality of these aspects. That it is done within the boundaries of the qualification structure, of the quality assurance system (examination bodies and inspection services) and of the financial frame.

Built into the WEB was an evaluation, the results of which had to be presented before the end of 2001. Therefore, in 1999 an evaluation committee was set up, composed of representatives of social partners, VET colleges and researchers. The Ministry of Education formulated a set of evaluative questions aiming to cover different key aspects of the reform. This steered the evaluation work, which was commissioned to eight different research teams. The role of these teams was to translate the evaluative questions into a set of useful measurable variables. In 2000 the evaluation teams started to work on different aspects, leading ultimately to a single report. The attributions of the different teams are presented in Table 3.4. The involvement of various research teams, with different specialities and backgrounds and different evaluation topics is an interesting feature of the evaluation of the WEB. As emphasised in Section 2.5.5, it enables capture of the variety of outcomes of a system's reform.

The evaluation Committee, using the various reports, formulated its own conclusions, which were presented in a common evaluation report in 2001, i.e. five years after implementation of the reform.

The main conclusion of the evaluation report was that although WEB has improved VET, there is still much inflexibility in the system for

⁽⁵⁴⁾ Nevertheless, the evaluation raised the issue of the efficacy of the tri-partite Danish system in terms of developing a dynamic and responsive VET system, rapidly accommodating changing professional profiles, and contributed towards an understanding of the reform within apprentice companies.

⁽⁵⁵⁾ This section is based on the detailed description of the evaluation of the Dutch VET reform (WEB) provided by Nieuwenhuis and Shapiro (2004: Chapter 6).

⁽⁵⁶⁾ WEB = *Wet Educatie en Beroepsonderwijs*. (Law on (adult) education and VET).

which institutional and organisational actors could be held responsible. Accountability is also badly organised. Furthermore, it appeared that the policy concept that drove the reform was not compatible with the requirements of the knowledge-based economy. WEB relies on the ability to prescribe and forecast qualification requirements, i.e. it requires a labour market with clear occupational assignments and well-established career pattern. The unstable and less predictable requirements of the knowledge-based economy require a different steering concept for VET. The reliance on flexibility of pathways and autonomy of colleges to define competence needs in collaboration with local companies should have been emphasised. However, the prescriptive nature of WEB had not incited the ROCs to organise responsiveness and flexibility at their level.

At the same time, the report indicated a shortage of expertise in ROCs, both at managerial and teacher level. The quality of assessments was especially poor. Recommendations made by the Committee put the emphasis on the need for greater autonomy and a more professional approach.

The report and its conclusions were received very sceptically by social partners (both employers' organisations and trade unions), which did not accept leaving the quality of VET mainly in the hands of the ROCs. Because of the controversial nature of the conclusions, the Ministry concluded that the evaluation was at the same time too early (five year is a too short time to evaluate labour-market success) and too late (to allow the institutions under assessment to benefit from it by correcting early enough what was not going well). A decision was taken not to design any new or corrective interventions, except for new regulations concerning the examination procedures (a new institute was established to prescribe how to organise assessments in colleges).

The approach followed in the Netherlands for the evaluation of the WEB was summative;

i.e. the research teams were supposed to take an *ex post* 'objective' picture of the state of the art on their respective subject. In a system like the Dutch one, where all political and social actors were involved in the formulation and implementation of the WEB, this approach made it difficult to identify failures, possible reasons and remedies *a posteriori*. This limited the value of evaluation results.

The use of the evaluation and its recommendations has been little and tends to greater institutional control (following the political debate) than the direction suggested (which was not well received in political circles). The policy learning that should have taken place as a result of the evaluation is not happening ⁽⁵⁷⁾.

A stronger involvement of social partners but also of practitioners in the design of the evaluation and the choice of evaluative questions might have reinforced the sense of ownership and then the acceptance of its results.

Absorbing the results of the evaluation of the reform means adjusting and correcting. The fact that the evaluation was conducted five years after large-/system-scale implementation of the reform might have created some resistance to change (again) from the actors of the system because new mechanisms are already in place and systemic changes have already occurred.

3.4.3. Learning from the Dutch and the Danish cases

'Whereas evaluations in VET have traditionally focused on outcomes and accountability, there is an increased interest from governments in new evaluation paradigms that can spur innovation and improve public policy through organisational learning processes.' (Nieuwenhuis and Shapiro, 2004, Chapter 3).

The steering and governance of change in VET is a subtle and complex enterprise. Policy-making (and adjustment) is probably well served by integrating formative evaluation in reforms. This would support implement-

⁽⁵⁷⁾ However, there is indication of some recent use of the evaluation results in some decision making, such as the socioeconomic council or the union of national colleges, for example.

Table 3.4. Evaluation of the Dutch reform, 'WEB', summary table

Evaluative questions	Main conclusions	Research teams ^(a)
<p>Links between VET and the production system</p> <ul style="list-style-type: none"> Is the qualification structure an adequate steering instrument for efficient labour supply? What are the supply and the quality of workplaces for learning? 	<p>Links between education system and labour market through the qualification structure are not (yet) implemented fully.</p> <p>Steering instruments implemented lead to a dispersed supply of courses while not delivering the expected flexibility.</p>	<p>Brandsma, de Jong, Karsten and van de Venne (University of Twente) Heijke, Borghans, Smit, Hövels and den Boer (University of Maastricht)</p>
<p>Response to social demand</p> <ul style="list-style-type: none"> Does the differentiation and quality of educational supply correspond to the individual educational demand? Is the assessment of prior knowledge and skills an appropriate input instrument? 	<p>'WEB' offers more possibilities for responding to social demands than are used by local actors. Autonomy tends to be restricted to lower levels of the system.</p>	<p>Doests, Westerhuis, Klarus, Meijers, Thomas, Vrieze, Dinjens, Neuvel and Pauwels (Cinop)</p>
<p>Quality of educational process and assessment procedures</p> <ul style="list-style-type: none"> What is the quality of the educational supply in terms of attractiveness of training, consistency between theory-practice-assessment, consistent translation of qualification structure into output? How well did external quality assurance mechanisms of assessment and examination procedures work? 	<p>Quality assurance has improved in the system although it restricts autonomy at institutional level. The system tends to be too regulated.</p>	<p>Nieuwenhuis, van Berkel, Jellema and Mulder (Stoas)</p>
<p>Efficiency</p> <ul style="list-style-type: none"> Has internal efficiency been reached: time/level ratio (amount of teaching and learning/educational level reached by the students), added-value (cost for reaching a specified level of low achiever students as compared to high achievers) and efficient internal trajectories? Has external efficiency been reached in terms of labour market position of people graduating from the system? 	<p>The system is more accessible. External efficiency and destinations have improved.</p> <p>Internal efficiency is difficult to estimate due to the lack of reliable data.</p>	<p>Van der Velden, Berken-bosch, de Bruijn, de Jong, Voncken, Geer-ligs and Lokman (University of Maastricht)</p>
<p>Synergies within the education system</p> <ul style="list-style-type: none"> What are the links with former and future (higher) education? What kinds of links exist between (initial) vocational education and adult education? 	<p>Cooperation between institutional actors has improved. The implementation of tailor-made individual trajectories for the participants is still in its infancy.</p>	<p>De Bruijn (Cinop)</p>
<p>Autonomy and quality assurance mechanisms at the level of the ROCs</p> <ul style="list-style-type: none"> What is the strategic power of ROCs to organise colleges as 'learning organisations'? What are the knowledge management capacities within ROCs? Was financial autonomy (ROCs receive a lump sum of money for VET and negotiate with the local communities for adult education, spending is at their discretion) implemented successfully? 	<p>Autonomy on the one side and regulations on the other tend to lead to tensions and contradictions. Although 'WEB aims at increasing autonomy at regional level, institutional regulation decreases it. Colleges do not use the full range of options available. Deregulation is recommended.</p>	<p>Karstanje, van Esch, van Ingen, Hoeven, Vermaulen, van de Venne (University of Amsterdam)</p>
<p>Implementation of legal aspects</p> <ul style="list-style-type: none"> How is the government coping with its new role (from steering and regulating to facilitating and controlling accountability)? What are the relations between different actors and stakeholders? What is level and quality of information to the public? What is the legal position of students? 	<p>The division of responsibilities between the different institutional actors is unclear. Although 'WEB' offers possibilities for an increased flexibility at local level this leads to a certain administrative burden of the ROCs'.</p>	<p>Leenknecht (University of Tilburg)</p>

^(a) All reports were published in Dutch by the Ministry of Education, Culture and Science in 2001.
Source: Nieuwenhuis and Shapiro (2004: Section 6.3) adapted by the author

ation, correct inadequacy or improve adequacy of a policy for initial goals and foster dialogue between the various actors, thus supporting a system change.

This is well illustrated by the evaluation of the Danish reform where evaluation is used as a learning exercise – a ‘reflective gathering of experience’ – after the pilot stage of the reform. Then it was used again immediately after the expansion of the reform to the entire system as a corrective and improvement mechanism. Evaluation is then part of the learning and changing process; the evaluators are involved, with system stakeholders, in the reform. In the Dutch case, the evaluation is conducted five years after large-scale implementation of the reform. It is summative, i.e. it is expected to assess and judge results. It is, therefore, used as an instrument for political control and its results are fed immediately into the political debate, thus generating reactions of a political nature. The potential for learning from the result is (partly) lost.

Summative evaluations are most useful when used after a pilot phase (before a new policy is implemented full-scale) and in collaboration with the main reform partners (institutions, teachers and trainers, individuals/participants). They provide an opportunity to identify success, failures and areas for improvement and can feed into reforms and policy formulation. In other words, their function is formative.

According to Lundvall (2000, cited by Nieuwenhuis and Shapiro, 2004: Chapter 3) knowledge construction and obsolescence is taking place at an accelerated pace. Therefore, it is not so important any longer to have knowledge *per se*. The key to success is learning (and forgetting). In the evaluation context, this means that new purposes for evaluation are emerging. Policy evaluation should also become an instrument of critical reflection and learning, in addition to measuring results and performance. Evaluations are given a new role in the learning process, both for policy-makers and actors in

the system (whether they are institutions or organisations, teachers or learners, etc.).

3.5. The forgotten army: teachers, trainers and VET professionals ⁽⁵⁸⁾

Teachers, trainers and other VET professionals are the ground agents of change and indispensable partners of reforms. Implementing a reform means not only providing teachers and trainers with appropriate training, it also means winning their commitment to the objectives of the reform. If the very people who are actually training people, managing schools or training centres, responsible for guidance services, etc., do not understand the purpose of a new policy, what it implies for them, and how it affects their role and functions, this new policy will face logical resistance and risk failure.

In contrast if the field actors are involved in, and convinced of, the necessity and relevance of the reform process through training, targeted information and participation in the evaluation, there is a higher probability that change will be implemented.

In addition, the kind of reform that affects the pedagogical role and functions and structure of VET (such as the qualification structure) must be transferred into initial and continuing training of VET professionals. This sounds obvious but examples show that it is not always the case.

Cedefop’s second research report (Descy and Tessaring, 2001a: 89-92) also underlined the key role that VET professionals (i.e. teachers and trainers, HRD managers, tutors at the workplace, etc.) might play in steering VET. Working hand in hand with education and training and VET professionals, while implementing new policy, is essential to introducing a coherent change strategy and ensuring internal consistency (Section 2.3). Involving education and training and VET professionals is, therefore, a strategic choice.

⁽⁵⁸⁾ This title was inspired by Frank Coffield, *The holes in the heart of current policies on lifelong learning*, intervention prepared for the conference *Policy, practice and partnership: getting to work on lifelong learning*, Cedefop, May 2003.

In the Danish Reform 2000, teachers and school managers were involved in the reform itself, in designing the evaluations, and were among the target groups for the surveys and interviews. This proved useful for identifying different practices, different roles (e.g. the role managers could play as catalysts of change), sources of difficulties (the stress felt by teachers because of their new duties) and successes (the contact teacher).

In Croatia, teachers and trainers have also become central to the reform process (Viertel et al., 2004, Section 4.5). When evaluating the Croatian VET system, the initial assignment given to the ETF was to analyse the situation of VET teachers' and trainers' training and to come up with proposals for an EU-funded development programme. After identifying some problems, the ETF decided to address the challenge of building upon internal links, while, at the same time, developing structures and actors. The key recommendation made was to decentralise strategic planning and budget execution and to increase the responsiveness of curricula. Implementation was to take place through strengthening the role of school headmasters and teachers

Decentralisation involved new educational management issues for school headmasters. Schools, apart from developing pedagogical innovations, were becoming responsible for strategic planning and budget execution (economising, looking for other sources of income, developing local partnerships, etc.). However, to increase responsiveness, decentralisation was also introduced in curriculum development and delivery. Schools had to be given more freedom in terms of specific content elements and learning methods, including a shift from teaching to learning-centred pedagogy. The school headmaster plays a role in promoting these approaches.

Teachers, for their part, had to be actively involved in translating the overall learning goals defined at central level into specific subject-matter, determining teaching plans and learning/teaching methods. They had to learn and use updated pedagogical equipment and to become familiar with a range of new

tasks and involved new processes and skills. Teams of teachers were also proposed to ensure better integration of the various subjects, including the integration of general and vocational subjects and of theory and practice. In this respect, the school headmaster had to play a supportive role.

A higher degree of decentralisation also means a greater need for quality assurance and accountability mechanisms. These had to become shared preoccupations for the whole school staff, led by the headmaster.

The required changes, therefore, could not be implemented without proper training (pre-service and in-service), both for the headmaster and teachers. It was suggested that all staff should be involved in a collective learning process instead of training teachers and headmasters separately on external courses. Learning was foreseen on-the-job, while changing day-to-day work processes, under the supervision of teacher-mentors. The teachers' trade unions had to be involved in this process. The ETF also recommended school development projects in partnership with universities, teacher training institutions and companies.

The specific set of recommendations made by the ETF, and included in the EU-funded development programme, were:

- (a) develop a 'change agent team' strategy by training a sufficient number of change agents whose role will be to introduce the new VET reform concepts and train teachers in schools;
- (b) develop mentors by selecting excellent teachers specialising in certain fields and training them;
- (c) design training programmes for school headmasters under the auspices of the Croatian school for headmasters;
- (d) design new methods for analysing training needs and a system to evaluate the relevance and effectiveness of training before, during and after the courses;
- (e) implement school development projects;
- (f) identify central teacher training institutions accompanying the development process and disseminating innovations.

In conclusion, ‘[...] any VET policy evaluation has to pay specific tribute, amongst others, to the professionalisation and continuous development of teachers and other education specialists, as they are the key to the success of any systemic VET reform effort.’ (Viertel et al., 2004)

Accomplishing change means influencing deeply embedded practices, routines, and beliefs, among teachers and trainers. It also means preventing resistance to change or attenuating its impact. Transforming education and training professionals into change agents contributes to achieving this. Because they are at the heart of change, education and training professionals are also a key partner in evaluation, whether formative or summative.

3.6. VET community programmes: case studies ⁽⁵⁹⁾

European Commission programmes in education and training aim to generate innovations and reforms in education and training systems of the Member States and of central and eastern European countries (CEEC). Evaluating these programmes includes, therefore, also aspect of system evaluation.

Evaluation of, capitalisation upon, and valorisation from programmes and projects are key concepts related to the desire to demonstrate value for money in EU programmes. They also relate to the need to make better use of experience, results and proposals emerging from the work undertaken.

Community programmes can act in two ways:

- (a) directly promoting the development or restructuring of VET in a state, as in the Phare programme;
- (b) indirectly generating impacts on VET development by creating a climate and special conditions through supporting projects and measures, as in the Leonardo da Vinci programme.

In both cases, planning and evaluation of impact on policies and valorisation of results are crucial to avoid community programmes becoming dead-end activities.

3.6.1. Evaluation of Phare VET programmes in the Czech Republic, Slovakia and Bulgaria ⁽⁶⁰⁾

Phare is an EU assistance programme to the CEEC in their process of transition to market economy and democratic societies. It supports preparation for EU membership. An important aspect of Phare assistance was the modernisation and reform of the CEEC VET systems. This was implemented in CEEC countries between 1993 and 1998.

The specific design and objectives of Phare VET programmes varied between countries according to the nature of reforms that were desired. However, some common objectives emerged:

- (a) to modernise and develop curricula for broadly defined occupations to increase responsiveness to the requirements of the economy;
- (b) to support institutional and policy development in VET;
- (c) to ensure education and training staff development.

In Phare VET programmes, planning foresaw pilot reforms in a limited number of VET schools before disseminating them to all schools. At the end of each programme, a team of international consultants carried out an external evaluation. The reports produced were rather summative in nature, assessing the results against stated objectives.

3.6.1.1. Czech Republic

In the Czech Republic, a Phare programme to support VET reform was implemented between 1994 and 1998. The Ministry of Education was the main stakeholder. The programme had defined objectives for the following subcomponents: curricula and pilot school development, pilot schools staff

⁽⁵⁹⁾ This section is based on Baumgartl et al. (2004); Lauterbach et al. (2004); Molsosa (2004).

⁽⁶⁰⁾ This first set of case studies was selected and discussed by Baumgartl et al. (2004: Section 2.1).

training, development of up-to-date learning materials, equipment in pilot schools, evaluation of programme results.

Results indicated that the investment made in the 19 pilot schools had a significant qualitative impact on their development. It exposed them to new pedagogical and management principles and initiated a quest for improvement and efficiency. However, positive results and good practice at pilot school level were not transferred to non-pilot schools. The evaluation indicates the lack of receptivity, interest and acknowledgement of results by the Ministry of Education that prevented the transfer of outcomes to the national level (despite dissemination activities carried out by the programme management unit). In addition, the involvement of third parties (e.g. social partners, stakeholders in adult training, tertiary education) was not fully achieved. This non-participation of third party stakeholders in the project impeded implementation of programme results on a larger scale. It also led to VET being approached in isolation from a wider lifelong learning perspective including tertiary and adult education as well as labour-market restructuring. Finally, the programme led to a (delayed) policy strategy. A School Act was prepared, taking into account the programme results, but a lack of political consensus led to several Parliamentary postponements. The attempt at reforming the Czech VET system lacked, therefore, both internal and external systemic consistency (see Section 2.3 for a discussion of these concepts).

The evaluation itself appeared to be of a formal nature. It assessed the programme against its stated objectives. The evaluators primarily studied the financial memorandum, the strategic plan, log frames, work programmes and reports. They also conducted interviews with managers at the National Training Fund (the project management unit), in the pilot schools and of stakeholders in the reform. In parallel, the management unit proceeded to a self-assessment. The evaluation did not consider impacts, nor did it undertake to understand principal pitfalls for the lack of acceptance of

content-related outcomes of the programme. There is little evidence of evaluating the value of products elaborated within the programme and no verification of its adequacy to the needs.

The evaluation was carried out exclusively by an international team of consultants (Section 2.5.5). In this respect, it did not contribute to enhancing the national experts' capacities in evaluation (although some local experts were involved).

3.6.1.2. *Slovakia*

The Phare VET programme in Slovakia, launched in 1994, intended to prepare new curricula for grades two to four in VET schools, to offer teacher training on the new methodologies and to provide new technological teaching equipment. It was piloted in 20 schools, before being transferred to the national system.

Unfortunately, the government that had accepted the programme left office a few weeks later and the following government did not approve the programme's general objectives. Social partners were, therefore, reluctant to participate in a project that had no government support. However, despite these obstacles, enthusiasm and effort were observed among individual teachers and school managers in the pilot schools.

No clear provision was made on how to transfer the new curricula to the VET system. Therefore, neither the curricula, and the developed teacher training, nor the experience accumulated in the pilot schools were used. Finally, one of the pilot schools even stopped using the then unofficial curricula and the question of the graduation of new curricula graduates was unresolved at the time of evaluation.

The evaluation assessed the achievements of the Phare programme and its relevance in developing VET to meet the needs of the market economy. It also considered the implementation of the programme. However, the lack of monitoring of the project by procedurally or politically responsible actors resulted in the absence of planning instruments, frameworks, targets, and

therefore of performance indicators. This hampered evaluation work.

Importance was attached to assigning a significant role to CEEC and Slovak experts to stimulate national discussions and transfer of results to VET policy-making. The international evaluation team was interdisciplinary. Extensive information collection, visits, surveys of pilot school and non-pilot school teachers, as well as a series of source material on the programme mandate and implementation, were used.

3.6.1.3. Bulgaria

The Phare VET programme discussed here was a direct continuation of a previous Phare programme that had introduced a general strategy for upgrading VET in Bulgaria. The initial programme focused on developing standards compatible with the EU for five professional branches as well as related practice and expertise. For both programmes, a pilot schools approach was adopted.

The second Bulgarian Phare VET programme, which started in 1996, was composed of several sub-programmes not inter related ⁽⁶¹⁾. This led to very diverse results, successes and difficulties. The phase of dissemination to the non-pilot schools followed a traditional pattern of publications, conferences and information-packages. Years after its completion, the programme had still not been widely implemented. Finally, new laws were adopted incorporating some aspects of the programme results.

The evaluation team included three Bulgarian evaluators along with three EU experts. The evaluation was based on:

- (a) analysis of the programme documentation and various implementation and final reports;
- (b) interviews and visits to the Ministry of Education and Labour, to pilot schools, regional authorities and institutions set up during the programmes;

- (c) questionnaires to programme recipients;
- (d) inputs from independent Bulgarian experts.

3.6.1.4. Some remarks on Phare VET programmes and their evaluation

The Phare programmes examined here, although they generated results and enthusiasm in pilot schools and initiated reflections and change at national level, presented two major drawbacks:

- (a) the absence of *ex ante* evaluation, i.e. of a phase where needs are defined based on a comprehensive socio-economic analysis, objectives are carefully identified and target and indicators of achievement set accordingly ⁽⁶²⁾;
- (b) the absence of a clear methodology for disseminating and mainstreaming the results achieved in pilot schools to the entire system. There was no consensus about goals and strategies for mainstreaming, agreement between the main actors, sufficient funding, clear objectives and an analysis of costs and benefits for the national system.

Phare VET programmes were carried out in parallel in several countries. They shared the same design (piloting approach), had the same aim of reforming VET and sometimes common broad objectives. In this context, exchange between programme managers, stakeholders and pilot schools across countries could have been a powerful tool of knowledge transfer.

Phare programme evaluations were targeted at each particular programme and were not trying to assess their links and interactions with other activities or with the VET system in general. In that sense, they lacked a system perspective, i.e. considering programmes and measures as relating to - and influencing - one another in addition to their individual effect (Section 4.4.3).

Because Phare evaluations were designed and managed externally and not by national

⁽⁶¹⁾ These subprogrammes were respectively dealing with upgrading VET; teacher career path; foreign language training; financial management for school education; development of science and technology; school drop-outs in general education; higher education evaluation and accreditation agency; a science park feasibility study.

⁽⁶²⁾ As discussed in Section 2.4.1, this stage is also the time to seek consultation and agreements between the various stakeholders.

actors, they were not perceived as tools contributing to policy planning and adjustment.

Finally, although the constitution of an international and interdisciplinary team is a positive feature (Section 2.5.5), the inclusion of an evaluator from the country is conducive to stimulating national discussions, developing national expertise in evaluation and contributing to enhancing the national evaluation culture.

3.6.2. The Leonardo da Vinci programme: selected case studies

Leonardo da Vinci I (LdV I, 1995-1999) was 'an action programme for implementation of a Community vocational training policy which supports and supplements the action of the Member States, while fully respecting the responsibility of the Member States for the content and organisation of vocational training and excluding any harmonisation of the laws and regulations of the Member States.' (Council of European Union, 1994).

LdV I took into account the experience of previous community actions (Commett, Eurotechnet, Force, Petra, Lingua). It provided a common framework of objectives along with a list of measures and priorities. The distinctive character of LdV I was determined by two basic requirements: transnationality⁽⁶³⁾ and innovation. The possibility for participation was offered to the CEEC.

In total, LdV I enabled 127 000 young people to undertake a period of training abroad; 77 000 organisations participated in various projects. The final report on the programme highlighted its complexity and management difficulties (Bainbridge et al., 2004: 29; EC, 2000). Requirements for evaluation were an integral part of LdV I 'at regular intervals on a partnership basis involving the Commission and the Member States.' (Council of European Union, 1994).

LdV is now in its second phase, 2000-2006 (LdV II). It⁽⁶⁴⁾ affirms the need to develop quality, innovation and the European

dimension in vocational training through transnational cooperation. Compared to LdV I, the programme is simpler and the number of objectives limited, but the same logic of intervention is kept. LdV II can be divided into five categories of measures:

- (a) mobility;
- (b) pilot projects;
- (c) language competences;
- (d) transnational networks;
- (e) reference material.

During its first two years, LdV II enabled 75 500 young people to participate in mobility measures and financed nearly 500 pilot projects (Bainbridge et al., 2004: 30; EC, 2002d).

As in the case of LdV I, requirement for evaluation is an integral part of the programme: 'The Commission shall regularly evaluate the implementation of this programme in cooperation with the Member States. [...] The main objective shall be the evaluation of the effectiveness and the impact of actions implemented in comparison with the objectives [of the programme]. [...] This evaluation shall also examine the complementarity between actions under this programme and those pursued under other relevant Community policies, instruments and actions.' (Council of European Union, 1999).

In this section we will look at two case studies:

- (a) the evaluation of the impact of LdV I in the Czech Republic;
- (b) the evaluation of the mobility measures in Germany (LdV I and II).

3.6.2.1. LdV I in the Czech Republic⁽⁶⁵⁾

Unlike the Member States, the CEEC entrance to the LdV programme was preceded by an *ex ante* study. This study focused on identifying national priorities in VET, the capacity of systems, the potential impact of the programme on policy and practices and on the formulation of specific proposals on participation (Handley et al., 1996).

⁽⁶³⁾ LdV I was the first programme to stipulate that the initiatives it finances should involve at least two Member States.

⁽⁶⁴⁾ Council Decision of 26 April 1999 (Council of European Union, 1999).

⁽⁶⁵⁾ This section is based on the background contribution of Baumgartl et al. (2004: Section 2.6).

The major expected impact of the programme was to forge a new mode of cooperation with Member State institutions on a partnership basis, so building qualitatively new relationships with the EU ⁽⁶⁶⁾, going beyond VET innovations.

Furthermore, in the CEEC, LdV I represented a major financial VET resource. It was, therefore, suggested in the *ex ante* study to select projects with a system and/or policy impact. However, LdV I, because of its transnational nature, lacked a procedure to give more weight to national priorities.

In 1999-2000 the countries were asked to produce 'national reports' on the implementation of LdV I. Because of their late accession to the programme and, therefore, the almost complete absence of finalised projects, the CEEC presented an interim report. Clear guidelines about the evaluation were not provided at national level, nor was it clearly stated whether the exercise should be performed internally or externally.

To evaluate LdV I, the Czech national coordination unit undertook two tasks:

- (a) the preparation of a report on the functioning of the programme, its results and influences on VET and employment in the period 1996-99, in order to issue recommendations for the next phase of the programme (NCU, NTF, 1999);
- (b) the commission of an external evaluation to a team of three independent experts ⁽⁶⁷⁾ in 2001, to valorise major results and best practice in the Czech LdV I (NA-NTF, 2002).

The first exercise was of a formative nature. It concluded that the major benefit of the programme was in placements and exchanges, especially in terms of enhancing the professional competence and foreign language skills of participants. Acquaintance with foreign VET systems, project design and management

skills, and transnational cooperation were considered important added-value. LdV I also brought about innovations, especially new educational products, and contributed to the harmonisation of Czech methods in VET with those in the EU.

The second evaluation was more summative and had two objectives:

- (a) to assess local, regional and national impact;
- (b) to identify best quality products and recommendations for valorisation ⁽⁶⁸⁾, focusing on the identification of projects potentially exploitable on a large-scale.

The evaluation focused on potential rather than real impact because many projects were just being finalised.

Comparing the programme objectives with the national priorities identified in major policy documents, the evaluators identified convergence of objectives and priorities, thus tending to demonstrate the contribution of LdV I to national developments.

Although many successful and useful products were identified, the evaluation could not estimate the possibility of mainstreaming results because it lacked analysis of systemic and legal barriers to such a large scale-transfer of innovations. The evaluation report, however, points out the high impact achieved by projects that attempted to meet the needs of a broad group of institutions. Indeed, limited impact was often caused, not by the unsuccessful realisation of the project itself, but by lack of external conditions for success. The involvement of various partners and institutions, and, therefore, more thorough reflection of varied conditions, increases the project's potential for a successful transfer.

One important conclusion of the evaluations is that international cooperation has a value in itself. In the Czech Republic, the national

⁽⁶⁶⁾ Rather than the usual donor-recipient relations existing between EU and CEEC.

⁽⁶⁷⁾ Each presenting expertise in one an area of the Czech LdV I programme: secondary and higher vocational education, tertiary education and labour market and employment.

⁽⁶⁸⁾ The concept of valorisation is particularly important in the context of education and training programmes. It is defined as 'the process of enhancing or optimising project outcomes through experimentation and exploitation with a view to increase their value or impact' (EC, 2002f, cited in Baumgartl et al., 2004: Introduction). It is also often used in the sense capitalisation on achievements or dissemination. Sometimes valorisation and evaluation are used as interchangeable concepts. Elsewhere, evaluation is considered as one aspect of valorisation. Valorisation activities can, therefore, turn out to be ambiguous in their purpose.

coordination unit adopted a strategy of involving as many people and as varied institutions as possible in LdV projects. This resulted in a great number of people who acquired experience in international cooperation and project management. This is attested by the high success of the Czech projects in more recent LdV calls.

3.6.2.2. *Mobility in LdV I and II: evaluation in Germany* ⁽⁶⁹⁾

The objective of the 'mobility' measure in LdV is to 'support the transnational mobility of people undergoing vocational training, especially young people, and those responsible for vocational training' (Council of European Union, 1999) ⁽⁷⁰⁾.

The basic objective is to strengthen the European dimension of initial and continuing vocational training, to encourage people to gain experience in activities involving theory and practice, and particularly work-linked training, to develop language skills and transnational contacts and exchanges of good practices for trainers and human resource managers.

The actual implementation takes place within three main areas of action, depending on the groups of beneficiaries involved:

- (a) transnational placement projects for students and recent graduates in initial vocational training and young employees;
- (b) transnational exchange projects for human resource managers, trainers, occupational guidance specialists trainers and mentors in language competences;
- (c) study visits for vocational training decision-makers, including social partners ⁽⁷¹⁾.

The *Institut für Wirtschafts- und Sozialforschung* (WSF, economic and social research institute) evaluated LdV I in Germany. It is also in charge of evaluating LdV II.

The evaluation of LdV I (Friedrich and Schumacher, 2000) drew on:

- (a) a brief analysis of trends and problems in the national education system;
- (b) an appraisal of various implementation reports and documents;
- (c) a summary of material and financial progress;
- (d) a systematic record of projects supported and a sample survey of 600 projects ⁽⁷²⁾;
- (e) consultation with actors involved and independent education experts (around 80);
- (f) structured interviews with programme coordinators ⁽⁷³⁾.

Evaluation of LdV II draws a fundamental distinction between mobility measures and innovative projects. The evaluation of the mobility measure is based on questionnaires sent to participants ⁽⁷⁴⁾ at the end of 2001 (1000) and at the end of 2002 (1000 + respondents from 2001). In 2003, 1000 participants plus respondents from the 1st and 2nd exercises were surveyed.

Some key findings in relation to mobility projects in Germany in LdV I are:

- (a) exchange measures involved approximately 30 000 people. This covers about 1/1000 of the total population of each of these groups (except for trainers). Therefore, the programme has no remarkable quantitative effect. However, as national bodies provide few opportunities for mobility measures, LdV has represented a useful improvement (Table 3.5);
- (b) systematic impact analysis of placement and exchange projects were not conducted by project organisers and insufficient time and resources were allocated to *ex post* evaluations. The limited evidence available, however, suggests that mobility projects are the most successful of all measures.

⁽⁶⁹⁾ This section is based on the background contribution of Lauterbach et al. (2004: Chapter 2).

⁽⁷⁰⁾ The objective and area of actions of the mobility measure do not substantially differ between LdV I and II. To present the measure briefly, we chose to refer to LdV II.

⁽⁷¹⁾ Programme managed by Cedefop on behalf of the Commission.

⁽⁷²⁾ 200 answers were received. The survey focused on central goals, dissemination measures undertaken and utilisation of project findings. Data of those projects received were weighted to relate to the total number of projects supported.

⁽⁷³⁾ This evaluation was carried out over a period of three months.

⁽⁷⁴⁾ School pupils, young people in initial VET, young employees.

Table 3.5. **Some figures for LdV I mobility projects in Germany (1995-99)**

Target group	Number of participants
Young people in initial VET	900 measures since 1995 42.3 % of all exchanges; 36.5 % of all projects
Young employees	continuing training courses: 6.2 % of all projects placement abroad: 15.3 % of all projects
Students or recent graduates	3.5 % of all projects
Trainers	21.4 % of all projects

Source: Hellwig et al., 2004, Section 2.3.1.0.

So far, evaluating the mobility measure of LdV II in Germany seems to be relying on a stronger design and on collecting evidence. This reflects the bigger concern in LdV II as a whole for monitoring and evaluation.

Data collected while evaluating LdV I, nevertheless, helped the implementation centres to improve the quality of the measure in the short-term. According to Hellwig et al. (2004: Section 2.4) evaluations in LdV have not usually taken into account the stated programme objectives. They focused on implementation and accountability monitoring, establishing whether a project was successfully organised, how many participants there were and whether they did well. 'The general assumption has been that a suitable number of successful individual measures with corresponding numbers of participants would automatically have a positive effect on the Europeanisation of vocational training in Germany. However, little empirical support for this hypothesis has been apparent to date' (Hellwig et al., 2004, Section 2.4). It is expected that the final report on the evaluation of LdV II will contain additional reference to goal attainment.

3.6.2.3. *Some recommendation for an evaluation of the LdV programmes*

LdV was created to promote better mutual understanding of VET in EU countries. Mobility measures serve that principal aim but their forms may differ quite substantially

LdV II, was established before the final evaluation of LdV I was completed (Box 3.7). Lessons were only partially taken on board in the priorities for the 1st call for proposals. Formally, it appears that in LdV II evaluations will be conducted more systematically than in LdV I.

LdV programmes are particularly difficult to evaluate because of the broad focus of goals pursued (although there has been a simplification in LdV II). Priority is given to the number and the diversity of beneficiaries rather than to the development of clear policy instruments.

Therefore, LdV generates a great number of projects, under various measures, for which it is difficult to draw a clear picture of impact. Consequently, evaluations tend to remain formal exercises, focused on implementation, financial data and number of projects and participants. They do not attempt to relate to the programme's initial objectives, or to measure the impact on the national system or at transnational level.

For mobility measures, Hellwig et al. (2004: Section 2.4) recommend using cluster evaluation. This method seems particularly suited to programmes with a common purpose but with different, relatively independent individual projects, having different strategies, contexts, etc. It seeks to investigate a group of projects, to identify shared themes through comparative observation. It follows what has happened within projects and tries to understand why. It is carried out in collaboration with project stakeholders: project managers, providers, evaluation commissioners and evaluators. Starting from the individual project level, this a bottom-up evaluation.

3.6.3. **Conclusions and ways forward for education and training Community programmes**

Baumgartl et al. (2004: Chapters 3 and 4) and Hellwig et al. (2004: Sections 6.2 and 6.3) make a series of recommendations based on their case studies analysis:

(a) only thorough *ex ante* evaluation and consultation, performed in parallel, create

Box 3.7. Developing policy-decision methods and accountability in the European Commission

Since 2000, the Commission has engaged in reform of its internal working methods and of its planning and decision-making procedures, thereby strengthening the role of evaluation.

Earlier, some steps had already been taken to streamline some policies and programmes by issuing various guidelines and launching initiatives on project management and evaluation (e.g. MEANS^(a); activities of DG Budget).

This led to a Communication on evaluation^(b), aimed at identifying and promoting good practice and strengthening methods of programme preparation. This communication recognises that although evaluations are increasingly carried out in most policy areas, further progress is needed, in particular in the quality and use of evaluation and its relevance for decision making.

Policy and community programmes in education and training

In education and training, the emphasis has been put on moving away from 'small' programmes or initiatives which have supported 'small' projects and networks. For example, the logic of intervention in LdV programmes (LdV I 1995-99 and LdV II 2000-06) is to support the cooperation of a large number of stakeholders in education and training at all levels^(c). In fact, education and training programmes try to conciliate two logics of intervention:

- a sprinkling logic, giving support to a great number of small projects of transnational nature to foster the European culture;
- a laboratory logic, targeted at decision-makers and promoting innovation strategies and policies in Member State systems.

The number and the diversity of beneficiaries of policy instruments and programmes leads to bringing together various priorities. Consequently, it is difficult to assess the achievement of initial objectives.

Evaluations of ET programmes carried out in the period 1997-2002 suggest the following conclusions:

- *ex ante* evaluations are often perceived as formal exercises, written after decisions have been taken;
- interim evaluations, concentrating on operational questions are useful for managers in terms of improving some procedures but they often lack an analysis of the relevance of the actions undertaken to attain initial aims;
- *ex post* evaluations are used to justify actions; they are often reports of execution;
- evaluations provide information on the results of programmes (number of projects, of participants, financial data, etc.) but there are difficulties in relating these results to the specific objectives of the programmes;
- all evaluations highlight the need for procedures for implementing and monitoring programmes;
- many evaluations highlight the lack of articulation with other community policies and with national and/or regional policies;
- programmes promote innovations and exchanges but are not real instruments of a community policy.

In addition, the use of evaluation results is often limited. First, timing tends to be bad. The pursuit of programmes is often decided before the results of *ex post* evaluation of preceding phases are available. Sometimes, as in the case of preparing

LdV II, interim evaluations are used. Second, evaluation stakeholders are many^(d) and this makes preparation of the evaluation, as well as use of results, complex. Third, in some Commission services, despite real progress, evaluations are still considered formal obligations to justify spending rather than tools for improvement. Managers are too busy with the direct management of programmes to design reflective evaluations and basic evaluations concepts are still not well known – even though the Commission has made specific training efforts in this respect. Finally, the budgets and delays necessary for evaluations are not sufficient. This leads to poor quality in offers received.

New orientations for the future

The reform of the Commission's working methods, initiated in 2000, is strengthening the role of evaluation as a management tool and as a support for planning and programming activities.

Ex ante evaluations of impact are now required by the Commission for preparing the annual policy strategy and the Commission work programme (EC, 2002e). The proposed method aims to structure decision making by systematically analysing:

- the problem addressed by the proposal concerned;
- the objective it pursues;
- the alternative options available to reach the objective;
- their likely impact;
- the respective advantages and disadvantages, including synergies and trade-offs^(e).

The aim of impact assessment is to improve the quality and coherence of policy-making.

In education and training, Community programmes are progressively being complemented by a political framework (Communication on lifelong learning, Lisbon strategy, objectives of education and training systems, etc.). The open method of coordination shows the determination of Member States to go further in this direction, by disseminating good practice and ensuring greater convergence in relation to some main EU objectives. From the Commission, this method requires a better capacity for animation, analysis and evaluation for it to be in a better position to make proposals for improving the effectiveness and efficiency of education and training. In addition, traditionally separated fields such as education and training are being integrated under the lifelong learning umbrella. These developments offer an opportunity to design clearer, more manageable and 'evaluable' future programmes (of which preparation has started in September 2002).

(a) MEANS: *méthodes d'évaluation d'actions de nature structurelle*.

(b) Communication to the Commission from Mrs Schreyer in agreement with Mr Kinnock and the President: *Focus on results: strengthening evaluation of Commission activities*. SEC (2000) 1051 of 26.7.2000.

(c) Although LdV II is simpler, with a limited number of objectives.

(d) E.g. in the case of LdV: Advisory committee for vocational training, LdV Committee, Employment and Education Councils, Commissions of the European Parliament, Directorate Generals for vocational training, management board of Cedefop, National Leonardo agencies, etc.

(e) The Communication on *Impact assessment* has been complemented by a Communication on *Evaluation standards and good practice* (Communication for the Commission from the President and Mrs Schreyer C(2002)5267 of 23.12.2002).

Source: Molsosa (2004).

- the conditions for a consensus on the objectives of the programme or the direction of the reform. If this is not done, authorities, institutions and actors at various levels do not identify themselves with the international project and do not take its results on board for further utilisation;
- (b) *ex ante* evaluation should be aimed at identifying themes and measures based on context analysis, researching main socioeconomic trends, labour-market requirements and institutional settings;
 - (c) national, regional and local VET stakeholders, including labour market institutions, need to define jointly the targets of the programme or the reform to increase adhesion and transparency (Sections 2.4.1 on implementing change and 2.4.2 on using building blocks).
 - (d) policies on projects in pilot schools should be supported nationally, regionally and locally to be relayed at systemic level after completion;
 - (e) a legal framework and institutions building are important conditions for achieving sustainability after a programme is completed;
 - (f) the programme design should include monitoring and continuing evaluation mechanisms. It should not be so complex that it causes difficulties in identifying appropriate and measurable achievement indicators. Time spent on project and indicator design is well-spent, as it will facilitate monitoring of implementation and final evaluation. It is an opportunity to translate broad and abstract goals into observable ones and if possible, measurable phenomena. Evaluation specialists should be involved in the design of programmes;
 - (g) dissemination is crucial. Different formats and presentations should be used to reach all stakeholders. They should be an integral part of the programme design and foreseen from the beginning;
 - (h) results of evaluations should also be widely disseminated and discussed, especially with those directly concerned and affected. The presentation of results should be targeted to the different recipients and the dissemination phase should be part of the evaluation design and financial schedule;
 - (i) evaluation should be understood as a development process and evaluation results should be disseminated to all countries involved in LdV, not only to the project partners;
 - (j) evaluation budgets are often too small to allow for a study of project achievements in detail and to ensure the dissemination of results to wake interest and prepare for further utilisation of the results;
 - (k) connecting a series of evaluations in a programme allows the stability of old findings to be verified and to medium and long-term causal chains illustrated. This requires not only access to previous evaluation and their data but presupposes their compatibility. It also requires that evaluation be incorporated in programme planning and annual operations;
 - (l) the international and interdisciplinary composition of evaluation teams is necessary to tackle the complexity of evaluation of programmes and their impact (see also Section 2.5.5 on interdisciplinarity). External evaluators need to be involved to complement national evaluation expertise;
 - (m) the inclusion of national experts, not only in international evaluation teams but also in all components of programme implementation, is desirable to ensure transfer of know-how, sustainability and capacity building in the country;
 - (n) all stakeholders of the programme should be involved and share a common understanding of evaluation goals, content and tasks. The risk of misunderstanding increases with the number of stakeholders involved, including international ones. This also requires, therefore, clarification of concepts and of the analytic methods used;
 - (o) Community education and training programmes are characterised by split responsibilities throughout the programme

cycle. Nobody has a full overview of what happens in a programme between the first idea behind it and the final recommendation. The huge number of people involved may lead to a loss of knowledge at each step as well as omissions and overlaps between the various stages. Sharing of responsibilities necessitates exchange of information, solid monitoring and reporting practices (as well as formats) and the assignment of a single project director (even at decentralised level) throughout the project duration.

3.7. Conclusions: evaluating and learning

According to Windham (2000, cited in Viertel et al., 2004), all evaluation work on education and training has three primary purposes:

- (a) to assess the nature of opportunities and constraints;
- (b) to assist governments, the private sector and individuals in establishing priorities, especially in a resource-constrained environment;
- (c) to specify the options for exploiting opportunities and dealing with constraints, to propose alternative strategies and goals and make concrete recommendations for policy reforms.

An evaluation in this sense is more than a reflection on the outcomes of a project or programme. It is primarily a tool providing advice and guidance as well as the basis for informed decision-making.

Evaluations should form an integral part of education and training policy, to contribute to improving decision-making. Its results should serve as a guide for decision-makers in choosing the most appropriate actions for specific objectives. It is, therefore, essential that all stakeholders, from learners to investors, understand the evaluation results and are in a position to grasp the challenges they present. To gain consensus, to maintain involvement from actors and to make policies viable in the long-run, it is crucial to assess the effectiveness and the utility of actions and

policies as well as to accompany their implementation (Pont and Werquin, 2004: Section 5).

One of the main conclusions of Viertel et al. (2004, Section 7.1) is that 'there is no holy grail in terms of conceptualisation or methodology related to VET policy evaluations. The engineer's toolbox is of limited use. [...] the only remedy is the evaluator's broad understanding of the essential components of the VET [system], of the relationships, of the fundamental logic between the system and its environment and of the strategic levers for change. This understanding develops only through many years of apprenticeship and first-hand experience with VET policy evaluation.' A community of practice for professional evaluators needs to develop.

There is also no standard or model of a 'good VET system' which the evaluator could use. There are only different solutions – which cannot be detached from historical, cultural and institutional contexts – to common challenges. Criteria for evaluation would depend on the scope and purpose of the evaluation exercise. Therefore, criteria should be subject to negotiations with the actors before the start of the evaluation exercise (Viertel et al., 2004, Section 6).

A systemic approach to VET evaluations should focus on analysing relationships and communication channels between institutions and actors in the systems and on identifying change-agents and internal and external (in)consistencies. This analysis should be based on the fundamental understanding that changes in one part of the system lead to changes in other parts and in the system as a whole. Each system is unique and no universal model can be applied.

Evaluation should balance aspects inside and outside of a system. Inside, evaluators are joint actors because they support a systemic change by organising critical reflection and learning. Outside, the evaluators' role is to assess the results of the changes implemented at a certain point. They are no longer interacting and must remain as neutral as possible.

A large-scale *ex post* evaluation carried out (years) after a new policy has been implemented is interesting for understanding the process of change and the results of policy at different levels. As a learning tool, it is very useful for social researchers. However, from a policy point of view, it can easily become an instrument for political control (sometimes even for others than those who have initiated the policy). Results are rarely used because their timing does not correspond with the policy cycle and evaluation reports often end up in drawers.

Guba and Lincoln (1989, cited in Nieuwenhuis and Shapiro, 2004), are quite critical of evaluators who see their task as verifying hypothesis and uncovering causal relationships between intervention and results/outcomes rationally and neutrally. Applying that positivist model requires a heavy battery of (semi-) experimental methods, statistical and econometric analysis. It helps to understand the relationship between variables under specific conditions. This kind of evaluation work (or rather this kind of research) is entirely justified, and of scientific interest but its value for social (in a broad sense) policy formulation, choices and design might be limited. This type of evaluation has to be seen as a complement of formative approaches, depending on the stage of reform.

The policy cycle is traditionally seen in different distinct stages:

- (a) decision-making;
- (b) implementation;
- (c) evaluation.

Lundvall and Borrás (1997) and many others call for a more fluid perspective of the policy process: a continuous transformation and evolution where no clear stages can be defined. It is artificial to see policy, and consequently evaluation, as a linear process. The three stages mentioned should be seen as a cycle where each stage is interacting with and dependent on the other. The initial phase of decision-making, policy design and planning is a phase of analysis and summative evaluation of previous and current policies, of their internal and external (in)consistencies. This initial analysis, in turn, greatly influences

what happens at the implementation stage, when choices become practices and when evaluation can be used as a corrective and constructivist learning tool. Finally, the last stage of *ex post* evaluation cannot be looked at in isolation from the others and ultimately it constitutes the first stage of a new policy. Following this logic, evaluation in itself is a continuous cycle that aims at accompanying and improving policy making, anchoring policy choices in social contexts and systems and ultimately leading to policy, institutional, organisational and individual learning.

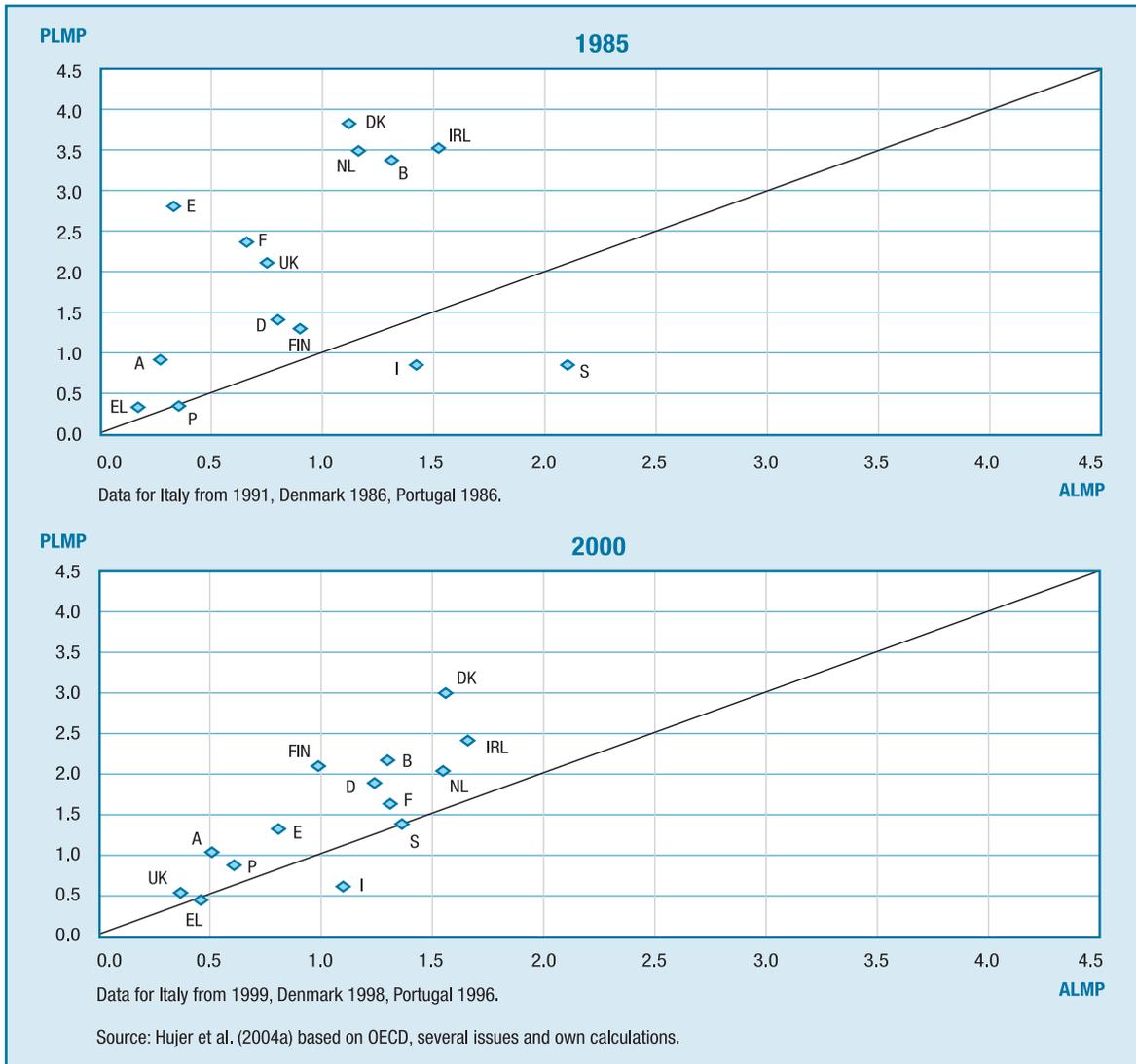
Evaluation is not just a matter of tracking progress against the desired policy goals; it is also a means of steering innovation in the intended direction, shared amongst stakeholders. Ideally, continuous feedback mechanisms from the field of action should allow for modification and adjustment of the policy. This (new) mode of policy formulation is better adapted to contemporary times of turbulent and constant changes. In contrast, the more static and traditional approach of summative evaluation, adapted to multi-annual periods of relative stability, no longer applies (Nieuwenhuis and Shapiro, 2004: Section 7.3).

4. Combating labour market exclusion: does training work?

4.1. Introduction

In this chapter, we will discuss policies and programmes, designed at national and EU levels, to combat unemployment, focusing on training measures. First, the relative importance of training as a form of active labour-market policies (ALMPs) will be discussed and its effectiveness reviewed. Second, we will consider policies and strategies developed at EU level to foster employment. Finally, we will conclude the chapter by discussing some drawbacks of common approaches to the evaluation ALMPs. We will argue that these approaches

Figure 3.13. **Spending on active and passive labour-market policies, EU15, 1985, 2000**
(% of GDP)



have to be complemented by other kinds of evaluation to draw a more complete picture of the benefits of training, thus putting evaluation at the service of policy-making.

4.2. Active labour-market policies

The persisting high level of unemployment during past decades (documented in Chapter 1) led to a growing awareness of the need to increase the effectiveness of employment policies. In particular, there is a belief in policy circles that active policies

should be introduced in addition to passive ones (which should include selective measures to help the heterogeneous group of the long-term unemployed).

The term 'passive measures' designates legal rights and entitlements that ensure that the unemployed receive unemployment and other related social benefits. Early retirement benefits are also covered under the term passive measures. ALMPs in contrast refers to more voluntary policies that foster reintegration into the labour market. They involve the deployment of various programmes targeted at specific groups and try to tackle specific problems affecting them.

According to the OECD (1996) ALMPs ‘seek to overcome difficulties of unemployed in finding a job and to improve the functioning of the labour market more generally’ (ibid. p. 7). Within ALMPs, significant resources are usually directed towards training, thus trying to improve the correspondence of unemployed people’s skills with those required in the labour market.

Since the 1980s, the trend in labour-market policy in Europe is to complement the distribution of unemployment and other benefits by active measures, the objective being to reintegrate people more quickly into the labour market. In various European countries, expenditure was progressively shifted between 1985 and 2000 and spending on passive and active measures is getting progressively equivalent.

For example, in 2000, Sweden and Greece spent equally on ALMPs and passive labour-market policies (PLMPs) while Italy was even spending more on active measures. The other EU countries still spent more money on passive measures although they seemed to be moving towards an equal spending pattern, if we compared to with the situation in 1985 (Hujer, 2004a; Figure 3.13). The OECD (2001b, based on LMP data, see Box 3.8) also reports that out of the total spent on labour-market programmes (LMPs), there is an increase of the share going towards active measures, i.e. it has grown from 31.7 % in 1985 to 37.7 % in 1998.

More recently, the OECD (2003a: 73) in its thematic review of adult learning, notes that common to all countries included in the review is the increased importance given to training and other active measures rather than the reliance on passive reception of unemployment benefits ⁽⁷⁵⁾.

There is, therefore, a common and strong belief in Europe that ALMPs should be implemented, in parallel to the distribution of benefits, to solve unemployment. Neverthe-

less, we will see that the ‘track record of many [active labour market] programmes is patchy in terms of achieving their stated objectives [and that] This has led many policy-makers to be wary in authorising large [or larger] spending increases on new or existing programmes’ (Martin, 1998) ⁽⁷⁶⁾ ⁽⁷⁷⁾. However, before discussing the evaluation and results of ALMPs, it is useful to have a closer look at their various forms and at the relative importance given to training measures in different countries.

The OECD (1996, 1998, 2001b) groups ALMPs into two main categories:

- (a) programmes that attempt, through developing employment-related skills, to make participants more productive and competitive in the labour market, thus increasing the responsiveness of labour supply and reducing the equilibrium rate of unemployment. The typical kinds of action in this category are training, retraining and job-search assistance;
- (b) employment programmes which attempt to stimulate job creation in both public and private sectors (including self-employment). The typical kinds of action in this category are subsidies to employment, direct job creation and aid to the unemployed starting enterprises.

However, several elements of these two kinds of action are often combined (e.g. job subsidy with training). In addition, these main categories are complemented by various actions aimed at enhancing the effectiveness of employment services (see Section 4.2.1), including developing job matching and guidance and counselling.

In this chapter we will mainly discuss evaluations of ALMPs that focus on developing employment-related skills, i.e. training and retraining (including job-search assistance). Other kinds of ALMPs, which aim to stimulate job creation, are only discussed

⁽⁷⁵⁾ For a discussion of the results of the thematic review on adult learning, see Section 2.2.1 and Annex 4. For a discussion of the method used by the OECD when undertaking policy review see Section 3.2.

⁽⁷⁶⁾ We will discuss results of evaluation of ALMPs, and especially training, in Section 4.2.2.

⁽⁷⁷⁾ Another possible reason why spending has not been shifted more from passive to active policies is rising unemployment. Passive income support is directly responsive to changes in unemployment rates (unless rules for eligibility for unemployment allowance are modified) while active labour-market programmes are more directly related to policy choices (OECD, 2001b: 23; Hujer, 2004a).

Figure 3.14. Public expenditures on ALMPs, EU15, 1999 (% of GDP)

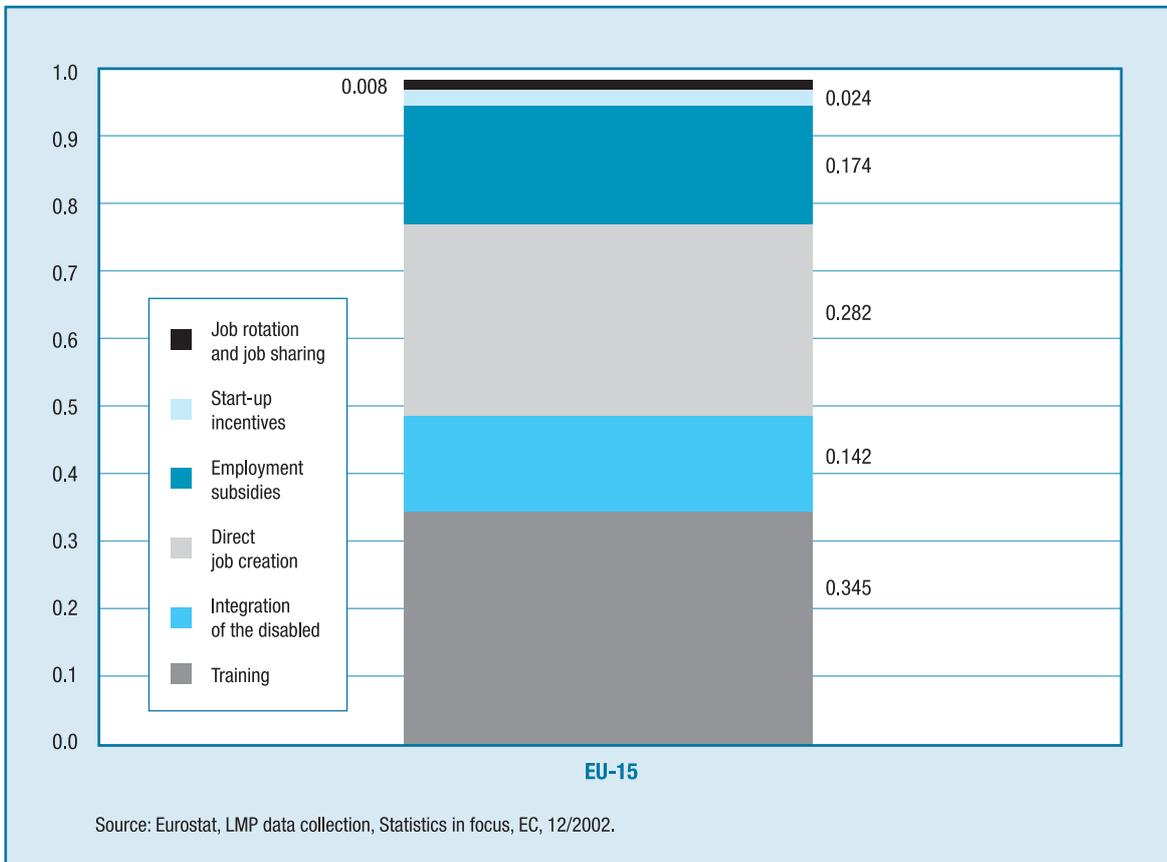


Table 3.6. Composition of active labour-market expenditure by type of action, EU15, 1999 (%)

	Training	Job rotation and job sharing	Employment incentives	Integration of the disabled	Direct job creation	Start-up incentives
EU15	35.4	0.8	17.9	14.6	28.9	2.4
B	15.1	9.8	15.2	11.4	48.3	0.2
DK	42	0	27.5	22.6	6.8	1
D	41.1	–	7.9	12.2	35	3.8
EL	35.1	–	30.2	26.7	–	8
E	27.7	0.6	42.4	9.5	14.2	5.6
F	30.6	–	21.3	9	38.9	0.1
IRE	25	–	6.6	3.6	58.8	6
I	35.5	0.4	42.9	0.7	17.3	3.2
NL	7	0	7.6	51	34.4	–
A	57.8	0	15.9	13.6	12	0.7
P	52.1	–	17.4	5.1	20.9	4.4
FIN	49.4	6.9	11.6	10.4	20.1	1.6
S	47.5	3.2	14.9	27.2	3.8	3.4
UK	51.6	–	6.2	26.4	15.3	0.5

NB: Luxembourg: not available.
Source: Eurostat, 2002.

Box 3.8. Data on labour-market policies

The OECD Labour market programme database

The OECD maintains a database on public expenditure on LMPs. It contains data that date back to 1985 and provides a basis for comparisons of expenditure between OECD member countries. The information covers all types of spending, including national, regional and local.

The LMP database includes information on the following measures:

- active measures: public employment services and administration, labour-market training, youth measures, subsidised employment, measures for the disabled;
- passive measures: unemployment compensation and early retirement due to labour-market reasons.

The information on training is limited to public spending and, therefore, excludes the private sector's spending on apprenticeship and other training. Similarly, training financed through payroll taxes is excluded.

The Eurostat Labour market policy data collection

The LMP data collection was designed as an instrument to follow up the targeted employment policies in EU countries resulting from the 1997 agreement to launch the European Employment Strategy. Eurostat developed it, in close cooperation with Directorate General Employment and Social Affairs, the EU Member States and the OECD. LMP data are made available annually.

The LMP data collection presents two main features: (1) detailed and comparable description of each labour-market policy measure; and (2) detailed information on participants.

It covers nine types of actions:

- active measures: training, job rotation and job sharing, employment incentives, integration of the disabled, direct job creation and start-up incentives;
- passive measures: provision of unemployment and early retirement benefits.

Public expenditure is distinguished by direct recipient (individuals, employers or service providers) and by the way the expenditure is disbursed (e.g. cash payment and foregone revenue). Participants are measured through three variables: stocks, entrants, exists.

Note: These categories also include hybrid measures for which it is difficult to make an account, e.g. when an employer is required to offer training in order to receive a subsidy.

Source: Walsh and Parsons, 2004; Eurostat, 2002.

marginally, in comparison with training.

The emphasis given to training and retraining within ALMPs can be estimated by looking at the distribution of spending on various ALMPs. Overall in the OECD countries, a quarter of the resources dedicated to active measures is spent on labour-market training (both for the unemployed and the employed at risk; OECD, 1998, 2001b, see Box 3.8). On average, in terms of % of GDP, training is the type of active measure in which the most money is invested in the EU15 (Figure 3.14; Eurostat, LMP database, see Box 3.8).

Considering the share of expenditure by type of action in individual EU15 Member States (Table 3.6), it appears that training is the most significant part of ALMP spending in Denmark, Germany, Greece, Austria, Portugal, Finland, Sweden and the UK. It is the second most important part of expenditure in Spain, France, Ireland and Italy.

Compared with EU countries, central and eastern European countries tend to spend less on training and retraining (Walsh and Parsons, 2004, Section 3.4). They favour other policies, such as job creation through public work and subsidised jobs. However, this might reflect the fact that these active policies are currently being introduced and that the full set of ALMP is not yet available ⁽⁷⁸⁾.

4.2.1. Improving effectiveness and efficiency of ALMPs

The rise of unemployment and the accompanying tightening of budgets lead to a growing concern about the effectiveness and efficiency of programmes and measures. Two strategies are commonly used to tackle this issue:

- (a) new management principles are introduced in employment services for administering and implementing ALMPs;
- (b) evaluation is used to assess whether the various policies implemented reach their objectives.

⁽⁷⁸⁾ For example, some of these countries spent a large proportion of their budget developing their employment services in the 1990s. Once the required level of service was established, they started focusing on other kinds of employment policies.

We will discuss both strategies in the coming sections.

4.2.1.1. *New management principles in employment services*

'Since growing budget deficits restrict the scope for spending increases, and since more spending does not necessarily mean better results, organisational reforms have been introduced to make policy formulation and implementation [...] more effective and efficient. Most of these reforms fall under the paradigm "new public management"' (Schmid, 1996: 747). According to this author the common features of the 'new public management' are:

- (a) management by objectives;
- (b) performance oriented budgeting;
- (c) strengthening competition between suppliers;
- (d) decentralisation of decision-making and implementation (Aucoin, 1990; Hood, 1991; cited in Schmid, 1996).

The OECD (1996, 1997) also underlines the critical role of the infrastructure running and monitoring the programmes regrouped under ALMPs. As public employment services are usually responsible for implementing these programmes, their effectiveness is an important aspect of successful ALMP ⁽⁷⁹⁾.

One of the new public management principles mentioned above refers to the necessity for strengthening competition between public services suppliers through the introduction of market forces. The use of market forces also stems from a desire to increase the freedom of choice for consumers ⁽⁸⁰⁾.

There are various ways of introducing market forces in the public sector:

- (a) privatisation (a public sector function is permanently moved into the private sector);
- (b) outsourcing (contracting a service outside the public sector);
- (c) quasi markets (a form of contracting out where the public provider can continue to

- act as one of the competing actors);
- (d) vouchers (enabling the client or customer to shop around between providers) ⁽⁸¹⁾ ⁽⁸²⁾.

Market forces are being introduced in many OECD countries in social policy fields such as health, education and housing. Four countries are considered as pioneers in their use of market forces in employment services (including training and reintegration programmes): Australia, the Netherlands, Sweden and the UK.

In Australia, three labour-market services were created: job matching (placement services for all registered job seekers), job search training (teaching job search skills to disadvantaged job seekers) and intensive assistance (individual counselling for the most disadvantaged job seekers). In the Netherlands, paying out benefits and associated tasks is still the responsibility of public services but from January 2002 reintegration programmes have been transferred to the private sector. The British authorities involve private parties in implementing various New Deal programmes in the different regions (targeted at young people, the long-term unemployed, single parents, etc.). Sweden has introduced market forces in organising the supply of training for job seekers. In 1986, the market for publicly financed training courses was opened up to public and private providers. In 1993, the government agency providing training courses (AMU Gruppen) was transformed into an independent state enterprise which has to compete with other players in the training market. The creation of open competition conditions is a key element in all countries.

The introduction of market mechanism is a part of the search for a more coherent benefit and activation system. The aims pursued with these reforms are a new service concept (Australia) or a more flexible supply of training (Sweden) or to develop outcome- and target-oriented services (Australia and the UK). In the

⁽⁷⁹⁾ Part 1, Chapter 3 of this report discusses in more detail the new management and accountability principles that accompany the running of a series of public services.

⁽⁸⁰⁾ This following text mainly draws on the paper from Struyven and Steurs (2004) on the evaluation of quasi-market reforms in employment and training services (background report).

⁽⁸¹⁾ For more information on training vouchers, see Cedefop's (Descy and Tessaring, 2001a: 66-67) and West et al. (2000).

⁽⁸²⁾ Except when the service is privatised, funding remains public.

Netherlands, market forces are aimed at reducing the number of people claiming disability benefits, while increasing the number of those leaving the benefit system for employment.

Struyven and Steurs (2004) evaluated these quasi-market reforms ⁽⁸³⁾. They conclude (mainly based on the Australian and Dutch experiences) that, for reintegration services, the market model is superior to that operating previously. However, this statement lacks empirical evidence and the authors conclude that the Australian case provides a mixed picture:

- (a) the market system leads to more cost-effective performance (services cost less, with a more or less similar net impact as in previous programmes);
- (b) responsiveness and quality appear rather sufficient;
- (c) the choices for job seekers are very limited and the level of services has decreased, especially in job-specific training;
- (d) short-term employment outcomes are pursued rather than training and education, which might, in the long-run, have greater benefits.

In the Netherlands, it is not clear yet how far the criteria of responsiveness and choice are met, neither do we have evidence of increased efficiency or equity. Struyven and Steurs conclude that the operational characteristics of the system matter but that they should be seen in the country specific institutional and labour-market context. They recommend that further evaluation research in this domain looks at the relationship between the outsourcing and market conditions on the one hand, and at the behaviour of both providers and job seekers on the other.

4.2.1.2. *Evaluating ALMPs*

Given the increasing importance of ALMPs in general and training in particular, one would expect that evaluating whether these meet their objectives, would be a common concern for governments. Surprisingly, it is not necessarily the case in Europe. Until recently, only a few European countries carried out rigorous evaluations (see Part 1, Section 3.1). In addition, despite the cost and political prominence of ALMPs, European policy-makers do not seem to pay enough attention to the results of evaluations.

Belgium, Germany, Ireland and the Nordic countries are inclined to carry out thorough evaluations, whereas in other countries there is less evidence of such practices. The weak evaluation culture ⁽⁸⁴⁾ is accompanied by a lack of appropriate data for evaluation. However, it is not always clear whether it is the lack of evaluation culture that impedes gathering of data or vice versa.

In contrast, a long-standing tradition of evaluating labour-market programmes exists in Canada and in the US (with the consequence that much of the literature emanates from these countries). 'A distrust deeply rooted in the North American society toward all government actions, combined with strong emphasis on the principle of individual responsibility, renders it virtually impossible to implement labour market programmes without an evaluation by independent experts.' (IZA, 2000: 2).

4.2.2. **ALMP evaluation results: what works?** ⁽⁸⁵⁾

Two main types of evaluative approaches to ALMPs can be distinguished ⁽⁸⁶⁾:

- (a) macroeconomic studies which seek to establish robust econometric relationships between key macroeconomic aggregates (such as employment or real wages) and various active policy measures;

⁽⁸³⁾ Struyven and Steurs, based on the theory of quasi-market of Le Grand and Bartlett (1993), assess the extent to which the models implemented in Australia and the Netherlands meet the success criteria for quasi-markets. For each criterion, a number of outcomes indicators were defined. In addition, process-related indicators relating to the functioning of quasi-market were added. These formed the framework for evaluating the reforms. For more details about national evaluations in Australia and the-Netherlands, the evaluations criteria and to how far the countries have met them, see the background report (Struyven and Steurs, 2004).

⁽⁸⁴⁾ Chapter 3 of Part 1 elaborates on the issue of evaluation cultures.

⁽⁸⁵⁾ Among other sources, this chapter is based on the contribution to the background reports from Walsh and Parsons (2004).

⁽⁸⁶⁾ See the discussion of macroeconomic and microeconomic evaluation in Part 2.

- (b) microeconomic evaluation of single labour-market programmes (Martin, 1998: 283-284).

In this section, we will mainly discuss the second type of evaluations (results and discussion on macroeconomic approaches will be presented in Part 4 of this report). We will sketch their main results and implications for future ALMP design, i.e. we will try to establish what works in ALMPs and what does not.

Before starting to discuss the results, first, a distinction should be made between different approaches that deal with the evaluation of 'single' programme:

- (a) those assessing the outcomes of programme participation on individual labour-market outcomes (mostly in terms of employment and earnings);
- (b) those investigating the net effect of programmes on total employment by estimating what would have happened in the absence of a programme (sometimes referred to as dead weight of a programme), substitution and displacement ⁽⁸⁷⁾ ⁽⁸⁸⁾.

Second, before starting to summarise the results of evaluations, it is also important to recall that most evaluations of ALMPs stem from the US, Canada and, more recently, Northern European countries (Belgium, Germany, Ireland, the UK and the Nordic countries). Evaluations are not widespread in central and southern European countries.

Table 3.7 summarises results obtained by various researchers in evaluating ALMPs. It is difficult to compare the relative impact of each type of measure to decide what is best among them but it is worth noting the following:

- (a) that special youth measures do not appear successful, unless this type of intervention is carried out early and is of broad scope;
- (b) job search assistance appears to help most of the unemployed, even if some conditions need to be in place;
- (c) subsidies to employment also seem effective, contrary to direct job creation.

Grubb and Ryan (1999: 80-92) conclude from their review of evaluation studies in the US that for disadvantaged adult men and women, gains in earnings were modest, while for disadvantaged youths, positive impact was absent and even some negative impact was observed. In Europe, training tends to increase the employability of disadvantaged participants but not their earnings. Again, the efficacy for young workers is dubious. In addition, even if training does have an effect on participant employment probability, the aggregate employment effect appears to be weak. '[...] public training often fails, particularly when it involves short, low-cost courses of remedial training and retraining, and when the criterion of success is a lasting gain in earning power, and not simply a short-term increase in employment rates for participants who continue to inhabit low-skilled labour markets. At the same time, public training can work when it sets its sights higher, aims at occupationally relevant needs in shortage labour market and takes training quality seriously.' (ibid.: 92).

After reviewing the results of microeconomic studies in selected European countries (partly summarised in Box 3.9 and Table 3.7), Hujer et al. (2004a) conclude that, in many instances, training programmes have positive effects; they found fewer examples of the positive effect of other types of ALMPs, especially job creation programmes.

On training measures as such, Fay (1996, cited in Walsh and Parsons, 2004, Section 4.1) concludes that while they are among the most expensive of ALMP options, expectations of their return for individuals should be modest. Nevertheless, small-scale, targeted programmes, that reflect the needs of both employers and job seekers, offer the best prospective outcomes.

Martin and Grubb (2001: 15) also conclude that, although training tends to be the most expensive active measure ⁽⁸⁹⁾, its efficacy is not obvious (at least in terms of traditional

⁽⁸⁷⁾ Displacement and substitution refers to mechanisms by which programme participants take the place in the labour market of the employed.

⁽⁸⁸⁾ In this chapter we will not go into detail of the various methods involved in each type of evaluation; this is done in Part 2.

⁽⁸⁹⁾ See also the indicators on the proportion of ALMP expenditure dedicated to training (Section 4.2, Figure 12, Table 6).

Box 3.9. Selected evaluations of ALMPs, with a focus on training, in some European countries

Evaluations of ALMPs carried out in Europe, are of a genuine 'European' nature. '[They present] a strong focus of labour market projects aimed at youth, an emphasis on employment effects rather than income effects as the outcomes of interest, and a complete exclusion of social experiments.' (IZA, 2000).

Sweden

Sweden is one of the most active European countries in conducting evaluation of ALMPs. This is due to the existence of the Institute for Labour-Market Policy Evaluation and of the availability of data going back to 1991. Various studies have, therefore, been carried out and some composite studies are also available. One of them (Calmfors et al., 2002) is of particular interest because it draws conclusions on the effects of training programmes in different periods. It concludes that, while in the early 1980s, positive effects were observed on employment and the earning potential of participants, in the 1990s these effects turned out to be insignificant and even negative. The authors suggest that this was partly due to the extension of the scale of programmes in the 1990s and the fact that the infrastructure was not able to cope. Another possible reason was the low demand for labour in this period of economic downturn.

Denmark

A study emanating from the Ministry of Labour (2000) provides a thorough appraisal of the Danish Employment Enhancement Programmes (EEPs) during the 1990s, relying on a number of studies carried out during that period on these programmes. The EEPs' key objective was to 'activate' the unemployed early in the unemployment period to prevent the drift into long-term unemployment. The effects of the EEPs can be summarised as follows:

- overall, positive employment effects for participants;
- while both public and private sector training programmes have significant positive effects, private training works better because participants tend to find work where they did their training;
- training programmes under EEPs reduce gross unemployment;
- by 1999, some 32 % of participants eventually got a job;
- the effects were less significant as the age of participants increased;
- the effect increased with higher levels of education among the participants.

A study by Jensen et al. (1999) looked not only at the inflows into regular employment but also into VET for participants in youth unemployment programmes. The effect on participation in further VET was positive, while there was no significant effect on employment.

Belgium

Nicaise (2000) carried out a meta-analysis of various evaluations of an employment-training programme for minimum income recipients in Flanders. This programme aimed at offering employment and training opportunities for the underprivileged, mostly the long-term unemployed. The author concluded that the programme had had a positive effect not only in terms of employment and earnings for the participants but also that other kinds of benefits were obtained, such as increased self-confidence, greater social contact, follow-up training and regularisation of social security rights. Furthermore, there were benefits for the municipality in terms of reduced workload for the welfare centres and lower incidences of crime and health problems.

Other studies found positive effects of in-firm training on the duration of employment (while external training and subsidised employment had no significant effect, Cockx et al., 1998) and of vocational training on outflow rate from unemployment (Cockx and Bardoulat, 2000).

Germany

Many evaluation studies in West Germany are based on the German Socio-Economic Panel (GSOEP-West) and use various econometric techniques. The results of several studies show negative or no significant effect of training on outcome variables such as unemployment duration (Hujer et al., 1998, 1999a, 1999b, 1999c; Hujer and Wellner, 2000b), employment probability and wages (Prey, 1997, 1999; Staat, 1997). Positive effects are detected for off-the-job training on reemployment probability (Pannenberg, 1995, 1996) and of public sector training on job search duration and wages (Staat, 1997).

In East Germany, studies are based either in the GSOEP East or on the East German Labour Market Monitor. There, positive effects of training measures have been more often detected. Training has positively influenced employment probability (Fitzenberger and Prey, 1997; Pannenberg, 1995, 1996), job security, earnings (Hübler, 1998; Fitzenberger and Prey, 1997) and the stability of employment (Kraus et al., 1998). Still a series of other studies tend to find no significant or even negative outcomes (Bergemann et al., 2000; Hujer and Wellner, 2000a; Lechner, 1999a, 2000; Fitzenberger and Prey, 1998, 2000).

Austria

Winter-Ebmer (2001) analysed the employment effects of job creation and training programmes organised by the Employment Services. These programmes were designed counteract staff reduction within the steel industry and only steelworkers were, therefore, eligible for participation. Results indicate positive employment effects primarily for unemployed older than 27 years, whereas for younger participants there were no effects.

France

Among the many public programmes available to young French workers, those offering some training raise the employment probability of participants. In addition, some studies show that this positive effect is stronger when young people lack prior qualifications (Bonnal et al., 1994). However this potential benefit is not widely available among disadvantaged youths because most of them are steered into job creation schemes in the public sector (without any training or with only a little; Balsan et al., 1996). Further evaluation of French vocational training and job creation programme (Bonnal et al., 1997) revealed that outflow rates from unemployment seem to be positively affected by vocational training (for the unemployed with a vocational education), while job creation schemes had no significant effect. Brodaty et al., 2001 found similar results using different evaluation methods.

Ireland

O'Connell and McGinnity (1997) analysed the effects of classroom training, on the job training and employment subsidy programmes on employment. Their results indicated that both training measures significantly raised employment rates, whereas for the employment subsidy programmes no significant effect could be found.

The UK

While the UK has a fairly consistent track record of programme evaluation, much of it is qualitative with only a few quasi-experimental studies. An example of the kind of approaches taken is the evaluation of the Prototype Employment Zones (PEZs) carried out by Haughton et al. (2000). The PEZs aim at generating greater flexibility and local innovation to help the adult unemployed (over 25 years old) return to the labour market. The evaluation was essentially a two-year longitudinal study tracking the development of the PEZs and involving fieldwork at various moments of the study. Although it was difficult to separate the relative effects from elements such as programme design, rules of engagement and prevailing labour-market conditions, the authors concluded that learning for work became the most popular option. It flourished partly because of the local provision developed, with new courses and greater flexibility.

Two other studies indicate the positive effect of training in terms of employment, while job creation programmes did not have any significant effect (Firth et al., 1999; Payne et al., 1996).

Note: We have not been able to identify the type of evaluation discussed in this box for other European countries. Evaluation carried out under the auspices of the ESF and the European Employment Strategy are discussed in Section 4.3.

Source: Walsh and Parsons (2004, Section 4.4); Hujer et al. (2004a), Grubb and Ryan, 1999.

outcome measures such as employment and earnings). 'Some programmes in Canada, Ireland, Sweden and the US have yielded low or even negative rates of return for participants when the estimated programme effects [...] are compared with the costs [...].' The most consistent positive results seem to be recorded for adult women while the most negative is for young people who dropped out of education and training: almost no training programme seem to work for them.

The results of selected studies carried out in Northern European countries (summarised in Box 3.9) tend to confirm that overall, it is difficult to draw a clear picture of the effectiveness of training from current evaluations' results. Effectiveness tends to vary across studies, according to external labour-market circumstances and across target groups.

Current studies suggest the following recommendations can be made for the design and implementation of training programmes with a view to increasing the effectiveness of active labour market (ALM) training ⁽⁹⁰⁾:

- (a) they should be closely targeted to specific groups and objectives (e.g. the long-term unemployed, unemployed youths, women returning to employment, illiterate adults, etc.);
- (b) they should allow for time to be dedicated to active job search (which also should be actively promoted);
- (c) they should be kept relatively small in scale and not be used as a large-scale solution to unemployment ⁽⁹¹⁾ ⁽⁹²⁾;
- (d) they should lead to qualifications or certificates not only recognised but also

⁽⁹⁰⁾ Based on the following literature: Fay (1996, cited in Walsh and Parsons, 2004), Martin and Grubb (2001: 16), Walsh and Parsons (2004), Hujer et al. (2004a), Grubb and Ryan (1999).

⁽⁹¹⁾ The extent to which programme coverage changes its effectiveness remains, according to Grubb and Ryan (1999: 42), an under-researched issue.

⁽⁹²⁾ The European Employment Strategy (EES) guidelines require that anyone who is unemployed for one year or more (six months in the case of young people) is offered training. This is a 'large-scale framework' within which specific, targeted and small-scale approaches should be implemented. A small programme may fail to benefit from economies of scale but a large programme may suffer substitution mechanisms in the labour market (Grubb and Ryan, 1999: 42).

- valued in the labour market (in this respect the involvement of employers in programme design and implementation seems to play a positive role);
- (e) they should establish strong links with local employers and provide work experience ⁽⁹³⁾;
- (f) when aimed at the young, they should be considered in association with general policies on education ⁽⁹⁴⁾.

The effects of training on employment and earnings tend to be modest but it is a very costly type of ALMP. In addition, it cannot be effective for all target groups, regions and episodes of the business cycle. Should it be concluded, in consequence, that training does not work well enough and that other kinds of ALMPs should be preferred? We will argue in Section 4.4 that establishing whether an intervention is successful or not, in terms of increasing employment probability and earnings for participants, is not enough.

First, there are other kinds of benefits of training that are usually neglected in the framework of ALMPs because they are not their primary goals. However, estimating benefits in personal terms, such as better self-esteem or motivation, and social aspects, such as better health or reduction in crime, should not be neglected ⁽⁹⁵⁾.

Second, in terms of informing policy-making, it is not enough to assess whether a programme worked or not. In order to design new policies, it is essential to identify ‘what worked?’ as well as ‘why it worked?’. What were the features of programme design or implementation that should be replicated in the future, and what should not? What were the specific external and internal conditions that

⁽⁹³⁾ Avoiding displacement or substitution.

⁽⁹⁴⁾ In addition to these, people participating in ALMP training should not feel that this situation is financially more interesting than being employed. However, and especially because training might postpone job search, the financial situation should not be worse than unemployment benefits and thereby decrease the motivation for training (Descy and Tessaring, 2001a: 355).

⁽⁹⁵⁾ Discussions in Part 4 about the material and non-material benefits of education training, at macro and at micro level, confirm that these kinds of effect exist. Perhaps it would be worth expanding the conception of the aims and the positive outcomes of training, even if it is used primarily to foster integration in the labour market.

Table 3.7. **Lessons from evaluation literature**

Appears to help	Appears not to help
Training/retraining	
Disadvantaged adults, especially women	Disadvantaged young people (?)
Long-term unemployed, in particular women	
Those laid off en masse	
	Young people
Men and women (young and middle-aged)	
Short-term unemployed	
Unemployed	
<i>of which:</i>	Formal classroom training
Women re-entrants	Prime-age men and older workers with low initial education
	On-the-job training
Women re-entrants; single mothers	Prime-age men (?)
Special youth measures (training, employment subsidies, direct	
	Disadvantaged youths
young people to find a job	young people to get higher pay
Job-search assistance (job clubs, individual counselling, etc.)	
Most unemployed but in particular, women and sole parents	
<i>of which:</i>	
Most adult unemployed	
Long-term unemployed; women re-entrants	
Men (below 40, relatively better educated)	
Direct job creation	
	Most adult unemployed

Source: This table is based on an original table from Martin and Grubb (2001), it has been completed and adapted by the authors.

on the effectiveness of ALMPs, composite studies

General observations

Source

- Type of training: remedial training of disadvantaged workers;
- mainstream programmes in the US: greatest gains in subsequent earning for adult women. Young register no gain or losses.
- in Europe, increase employability but tends to leave earnings unchanged;
- the picture does not seem as negative concerning the effect on youth in Europe as it is in the US.

Grubb and Ryan (1999)

Small positive effect on employment and earnings, which depends to a great extent on the business cycle.

Dar and Tzannatos (1999) ⁽¹⁾

- Has a high dead-weight effect;
- most successful if on small scale and targeted towards the most vulnerable groups.

Dar and Tzannatos (1999) ⁽¹⁾

- Did no better than control groups in terms of improving employment probability and earnings;
- social rates of return tended to be negative.

Dar and Tzannatos (1999) ⁽¹⁾

The impact is higher for women than men. The impact is lower for older participants than the young. Those with primary and secondary education benefit more than those with post-secondary education.

Fretwell et al. (1999) ⁽²⁾

Fretwell et al. (1999) ⁽²⁾

In central and eastern Europe countries. Albeit with some limitations imposed by the general state of the labour market. In these countries, training emerges as a relatively low-cost measure.

Walsh and Parsons (2004)

- Important that courses signal strong labour-market relevance or signal high quality to employers;
- keep programmes relatively small in scale.

Martin and Grubb (2001) ⁽³⁾

- (?) Because some programmes gave positive results and others not;
- must directly meet labour-market needs. Hence, need to establish strong links with local employers, however increasing the risk of displacement.

Martin and Grubb (2001)

job creation measures

- Effective programmes need to combine an appropriate and integrated mix of education, occupational skills, work-based learning and supportive services to young people and their families;
- early and sustained interventions are likely to be most effective;
- need to deal with negative attitudes to work on the part the young. Adult mentors can help.

Martin and Grubb (2001)

- Insertion programmes;
- evidence studies in France and in the UK;
- gains for young people result in displacement for other workers.

Grubb and Ryan (1999)

Must be combined with increased monitoring of the job search behaviour of the unemployed and enforcement of work tests

Martin and Grubb (2001)

Re-employment bonuses

Requires careful monitoring and controls on both recipients and their former employers.

Martin and Grubb (2001)

Subsidies to employment

Require careful targeting and adequate control to maximise net employment gains and social benefits. Some substitution mechanisms might reduce net employment gains.

Martin and Grubb (2001)

Aid to unemployed starting enterprises

Only works for a small subset of the population.

Martin and Grubb (2001)

- Typically provides few long-term benefits and principle of additionality usually implies low marginal-product jobs;
- no significant or even negative effects in France, Switzerland, Sweden and the UK

Martin and Grubb (2001)
Hujer et al. (2004a) ⁽⁴⁾

⁽¹⁾ Cited in Walsh and Parsons (2004)

⁽²⁾ Based on a review of evaluations up to 1999 in four transitional countries: Czech Republic, Hungary, Poland and Turkey

⁽³⁾ Martin and Grubb (2001) based their finding on : US Department of Labor (1995), Fay (1996), Friedlander et al. (1997), Grubb (1995), HRDC (1997), Lerman (1997) and OECD (1993)

⁽⁴⁾ Based on Laive et al. (2000), Gerfin and Lechner (2000), Bonnal et al. (1997), Brodaty et al. (2001), Firth et al. (1999); Payne et al. (1996) and Larsson, (2000).

Table 3.8. **Selected results from national reviews of Employment policies – EES, impact evaluation (focus Pillar I: employability, in particular training measures)**

Country	Type of measures	Review methodology	Main results
Spain	<p>Training for the unemployed:</p> <ul style="list-style-type: none"> • design and delivery of training are decentralised to regional authorities; • main target groups: short-term unemployed youth and adults. 	<p>Econometric analysis of the impact of measures for participants:</p> <ul style="list-style-type: none"> • chance of having a job at the time of the analysis (hiring rate); • chance of having a job at one point during the observation period (employability rate). • estimation techniques for comparing the treatment group against a control group. 	<p>Training had positive effects for most participants and significantly improved the employability of certain sub-groups (in particular disadvantaged youths, disabled persons and ex-convicts).</p> <p>Employability rate offers a better indicator of the longer-term job prospects of participants.</p>
The Netherlands	<p>Overview of a wide-range of training measures.</p>	<p>Drawing on previous evaluation (in particular De Koning, 1998).</p>	<p>Effects of participation in training programme changed significantly throughout the 1990s. Until 1993, there were significant positive effects, especially for long-term unemployment. After 1997, no significant effect of programme participation can be detected.</p>
Ireland	<p>ALMPs under the National Employment Action Plan.</p> <p>Support to the unemployed at an early stage of unemployment period mainly by providing a job, following with guidance and counselling (within this process a place on a training programme might be offered).</p>	<p>Administrative data and consultations with key players of social and labour market fields.</p>	<p>Although the report reaffirms commitment to such measures to help alleviate social exclusion and labour shortages, it also questions whether they always work. It suggests that the probability of leaving unemployment is rather dependent on the individual's capacity to compete in the labour market than on the support provided through ALMP.</p>
The UK	<p>New Deal Programme: targeted at young people (18-24) once they reach six months of unemployment and those over 25 once they have been unemployed for 18 months.</p> <p>New Deal 50 Plus is targeted at those unemployed aged 50 and over, who are eligible after six months of unemployment.</p>	<p>Administrative data.</p>	<p>Reasonable signs of success with 57 % of sample in work (the majority from the younger age groups, have shorter unemployment spell and are women).</p>

Country	Type of measures	Review methodology	Main results
Austria	<p>Training programmes upgrading the skills of the unemployed and responding directly to employers' needs.</p> <p>Qualification enhancement programme through public employment service.</p>	<p>Micro-analysis of programme participants and non-participants, adjusting the gross-findings with econometric techniques to calculate estimates of net impact.</p> <p>Estimation of macroeconomic effects of policies.</p>	<p>Positive macro impact of participating in training programmes, when upgrading of skills of the unemployed corresponds directly to employer needs. Increase of aggregate employment by over 9 800 each year. Contribution of 0.32 % to GDP growth.</p> <p>Earnings potential of participants in the qualification enhancement programme increased by EUR 2 870 per annum (in comparison with control group).</p>
Finland	<p>Thorough appraisal of all four pillars.</p>	<p>Administrative data combined with other sources such as the Labour Force Survey.</p> <p>Various manipulations (including quasi-experimental techniques).</p>	<p>Overall employment effect of training measures were modest for the individuals concerned.</p> <p>Labour-market training has a weak net impact on employment probability of participants (outcomes slightly better for those with lower than average employability at the start of the programme).</p> <p>Interview-based survey shows considerable social benefits for programme participants. The study concludes that training designed for immediate employment should be responsive to the economic cycle, being reduced when fewer jobs are available. Training that helps to increase skills should, on the contrary, respond inversely to the economic cycle.</p>

NB: Walsh and Parsons did not consider reports from the other Member States because of their lack of specific coverage of ALMPs and attention to detail.
Source: Walsh and Parsons, 2004 based on national reports prepared in the context of the 2002 Impact evaluation of the EES (http://europa.eu.int/comm/employment_social/employment_strategy/impact_en.htm).

allowed the intervention to bear fruit? Was the programme implemented as expected? Was the logic of intervention appropriate? These are examples of questions which common forms of evaluation of ALMP fail to answer fully.

In Section 4.4 we go into more detail on these issues and we make recommendations for evaluation design. Before that, the next section completes the picture on employment policies by presenting the European Employment Strategy (EES) and how it contributes to creating an evaluation culture in Europe. We will also present some results of evaluations carried in the context of the European Social Fund (ESF).

4.3. Fighting unemployment at EU level

European countries do not all share the same evaluation tradition. However, the need to carry out evaluation of ESF projects as well as of reviewing progress in the framework of the EES led all EU Member States to carry out some form of evaluation or monitoring.

4.3.1. The European employment strategy (EES)

Concerns about the high level of unemployment in the EU led to the EES, a framework for a more coordinated policy response. Training is an important aspect of this strategy. This section summarises the evolution of the EES from its start in 1997 until now and presents selected

Box 3.10. EES Peer review programme on ALMPs

The dissemination of good practice among Member States has the overall aim of enhancing transferability and mutual learning as well as promoting a greater convergence towards the main EU goals. It is part of the EES.

The European Commission (DG Employment and Social Affairs) has launched a programme in close consultation with the Employment and Labour Market Committee, to promote the identification and exchange of good practices in ALMPs throughout the EU. To this end, peer reviews have been conducted since 1999. The participants of a peer review are government representatives, independent labour-market experts and representatives from the European Commission.

The objectives of the peer review programme are:

- to identify, evaluate and disseminate good ALMP practices;
- to assess whether and how good practices can be effectively transferred to other Member States;
- to follow-up and implement elsewhere the ideas and objectives of the EES;
- to develop criteria for selecting and reviewing good practices.

Selection of good ALMP practices

Good practices to be reviewed are selected by Member States from those published in the joint employment reports or the national action plans for Employment. They are based on existing evaluation or early monitoring data and thematically linked with the pillars and guidelines of the EES.

Each peer review is hosted by the Member State in which the selected practice has taken place. It is attended by a group of peer countries with a special interest in the experience and in the transfer of the policy. Since 2002, ÖSB-Unternehmensberatung and The Institute for Employment Studies (IES) have assisted the European Commission in organising the reviews and disseminating their results. In 2000-01 the reviews were organised by ÖSB-Unternehmensberatung and INBAS, in 1999 by NEI.

Source: <http://peerreview.almp.org/en/principles.html>

results of its impact evaluation.

The decision to pursue a common EES was taken at the Luxembourg summit in 1997, recognising the need to act collectively at EU level. Its implementation started in 1998. The

aim has been to enhance labour-market policy while providing a common European framework for dealing with unemployment.

The EES is a form of 'management by objectives'. It is based on peer review and benchmarking policy and practice between Member States ⁽⁹⁶⁾. The Council, following a proposal from the Commission, decides on employment guidelines each year. These guidelines have to be transformed into national policies and reported on in national action plans (NAPs). NAPs are then assessed with a view to setting the next annual guidelines. Since 2000, in addition to the employment guidelines, the Commission issues specific recommendations to Member States. Therefore, embedded in the EES is the requirement they review their labour-market policies on a regular basis. Additionally, a peer review programme is under way to foster the dissemination of good practice (Box 3.10). As such, the EES contributes to the development and establishment of evaluations and mutual learning culture in EU Member States.

In 2001/2002, Member States carried out a thorough review covering the three first years of the EES (from 1998 to 2001). This provided a useful insight into the approach to active employment policies, as is summarised in Table 3.8, and served as a basis for the 2002 *Impact evaluation of the EES* (EC, 2002c).

This Impact Evaluation concluded that the comprehensive approach of the EES generally strengthened national employment policy coherence and framework. The political commitment of Member States is reflected on in the NAPs and the practice of setting targets towards reducing unemployment is undertaken by some of them. However, there is still a widespread reluctance to set targets and to make visible budgetary issues. The EES has also fostered agreements on new common paradigms such as lifelong learning (ibid.: 9-10). Member States have increasingly used VET as a preventive and active labour-market tool to promote employability, adaptability and competitiveness (Bainbridge et al., 2004: 32 ff.). The target set in the Employment Guid-

⁽⁹⁶⁾ See also Sections 2.5.2 and 3.3 on benchmarking in the EU.

Table 3.9. Summary of main findings of ESF evaluations, Objectives 1, 3 and 4

Objective 1: to develop regions which are currently underdeveloped
<ul style="list-style-type: none"> • net impact was greater when programmes concentrated on the most disadvantaged groups than when they were targeted to all the unemployed; • in some cases, combining measures increased their net impact; • programmes offering work experience achieve better results when combined with some form of skills training.
Objective 3: to tackle long-term unemployment, promote equal opportunities, improve lifelong learning, encourage entrepreneurship and adaptability and improve the role of women in the workforce
<ul style="list-style-type: none"> • overall, gross employment impact ranged from 30 to 80 %; • clients showed high levels of satisfaction with their programmes; • the chances of finding work after a programme depended less on the beneficiaries' personal characteristics than on the availability of jobs in the labour market and the type of project.
Objective 4: to improve the qualifications and prospects of all those in employment
<ul style="list-style-type: none"> • some evaluations suggested that employers benefited more than workers from the supported activities; • future skills needs were generally not identified and taken into account into project design; • some of the training supported through projects was considered too general or, in some cases, not sufficiently transferable between employers.

Source: Walsh and Parsons (2004) on the basis of European Commission (2001b).

elines, to offer active measures such as training and subsidised employment to 20 % of the unemployed, has generally been reached. In addition, training measures prove to be effective for particular target groups, such as women re-entering the labour market and educated immigrants.

On the basis of the 2002 impact evaluation, the European Commission (2003b) made a number of recommendations on the future of the EES that were implemented in the 2003 guidelines. They are articulated around three overarching objectives: full employment⁽⁹⁷⁾, quality and productivity at work and strengthening social cohesion and inclusion. They focus on 10 key priorities, supported by a series of actions and, when appropriate, benchmarks⁽⁹⁸⁾ (Bainbridge et al., 2004: 34). Mutual learning and a culture of evaluation should continue to be fostered and developed with the support of three kinds of tool:

(a) the peer review programme, which should be considered a key component of the

EES in order to facilitate the transfer of know-how between existing and new Member States (Box 3.10);

- (b) a continuous evaluation programme to be developed at EU and national levels;
- (c) indicators, which should play a major role in the evaluation of national employment policies and of EU labour-market performance.

4.3.2. Evaluations carried out for the European Social Fund

The European Social Fund (ESF) provides funding on a major scale for programmes which develop or regenerate people's employability. This task focuses on providing citizens with appropriate working skills and developing their social interaction skills, thereby improving their self-confidence and adaptability in the job market.

The ESF purpose is to support measures which aim to prevent and combat unemployment, develop human resources and foster integration into the labour market, to promote

⁽⁹⁷⁾ The full employment objective refers specifically to the Lisbon and Stockholm targets:

(a) an average employment rate of 67 % in 2005 and 70 % in 2010; (b) an employment rate for women of 57 % in 2005 and 60 % in 2010; (c) an employment rate of 50 % of older workers (55-64) in 2010.

⁽⁹⁸⁾ See discussion of the benchmarking approach in Section 2.5.2.

a high level of employment, equal opportunities for men and women, sustainable development and economic and social cohesion. In particular, it must assist measures taken in line with the European strategy and guidelines on employment.

The ESF channels its support into strategic long-term programmes which help regions across Europe, particularly those lagging behind, to upgrade and modernise workforce skills and to foster entrepreneurial initiative. This encourages domestic and foreign investment in the regions, helping them to achieve greater economic competitiveness and prosperity.

Programmes are planned by Member States with the European Commission and then implemented through a wide range of provider organisations in both the public and private sectors ⁽⁹⁹⁾.

The funding requirements and the implementation of the ESF also play a role in developing an evaluation culture throughout the EU because the ESF requires evaluation to be carried out at three levels:

- (a) project level: each funded project is required to have an explicit monitoring and evaluation strategy;
- (b) national level: in the 1990s Member States were involved in mid-term and final evaluations of their national ESF programmes;
- (c) European level: the European Commission, drawing on the national reports, indicates EU-wide trends, of which Table 3.9 summarises some findings.

From the EU-wide report, it appears that targeted policies always work better than non-targeted ones and that linking training to labour market and employer needs is always likely to improve their impact.

4.3.3. Conclusions

Although the EES and the ESF contribute towards raising the profile and the use of evaluations, benchmarking and peer reviews, there are still substantial differences across EU

countries in terms of using rigorous evaluation methods.

Overall, evaluations carried out in the framework of EU-supported policies display the same pictures as the more traditional evaluation of ALMPs: there is no consistent indication of the effectiveness of training when it is used as a means geared at getting people back into employment. Training seems to have a positive effect but it tends to be modest and linked to the prevailing labour market conditions. Training needs to be targeted and it works better if employers are involved or if their needs are directly taken into account. In difficult labour market conditions, training might not be sufficient as a measure to get people back into employment.

4.4. Recommendations for ALM training evaluation

Two main conclusions emerge from the examination of the various attempts to evaluate the effectiveness of ALM training:

- (a) current evaluation results do not provide a clear picture of the effectiveness of training; it tends to vary across programmes and target groups as well as across studies;
- (b) exogenous circumstances affect the relative effectiveness of training considerably.

Local circumstances, including prevailing labour market and administrative conditions, also limit the degree of generality and transferability of evaluation results. As a consequence, the recommendations that can be drawn from reviewing evaluation work lead to 'identifying common denominators that hint best practices' (Walsh and Parsons, 2004, Section 5.2).

Additional factors render evaluation work highly complex and tend to limit the usability of evaluation results ⁽¹⁰⁰⁾:

- (a) evaluation results are highly dependent on the methods and data used. Examples

⁽⁹⁹⁾ This brief description of the ESF is based on the DG Employment and Social affairs ESF 2000-2006 web site: http://europa.eu.int/comm/employment_social/esf2000/introduction-en.htm#key.

⁽¹⁰⁰⁾ Partly based on Martin and Grubb (2001: 10-13).

show that for one specific programme, different evaluations can draw different conclusions (Hujer et al., 2004a);

- (b) usually evaluations of ALMPs focus strictly on earnings and employment as positive outcomes. Accounting for non-intended results (including negative ones) is also necessary, even though it generally falls outside the scope of evaluations. In addition, a programme could also bring other non material benefits, such as reduction of crime, better health, or even well-being or changes in attitudes, for participant and/or for the society as a whole. It could also influence the way the labour market or businesses operate. Finally, the fact that training might lead to further training (and not to employment immediately) is, in most cases, not considered (and sometimes considered as a negative outcome because it delays the return to employment);
- (c) it is not only policy-makers and evaluation commissioners, who want quick results; the evaluation methods themselves incline evaluators towards measuring short-term outcomes ⁽¹⁰¹⁾. Evaluation budget constraints also play a role in this respect. In consequence, the vast majority of evaluations cover at best a period of one to two years after people's participation in a programme. This is probably too short a period for a full assessment of private and social returns;
- (d) ALM programmes tend to be small-scale. Hence, even if a small programme has significant positive results for participants, it does not mean that it would remain cost-effective and produce the same kind of positive outcomes – without generating adverse effects – as if implemented on a larger-scale (in terms of number of participants or in geographical terms). This remains an under-researched issue but it seems that the content of programmes

matters more to their effectiveness than their scale of operation (Grubb and Ryan, 1999: 42);

- (e) answering the question 'what works' is not enough, in particular if one aims at informing policy-making. It is also necessary to find out why a programme works for one group and not for another? What are the characteristics of a programme design that led to (the absence of) positive outcomes? What combination of employment services is likely to bring best results? What are the external circumstances that influence a programme's effectiveness? The most common forms of ALMP evaluations are not tackling these issues;
- (f) ALMPs change according to specific needs and local contexts, various types of programmes are implemented in parallel and the mix of programmes is continuously changing. As the passive and active measures seem to be used more frequently as one complementary to the other, the work of the evaluator is likely to be further complicated. This renders the task of the evaluator very complex and is confusing for potential programme 'clients' (firms and people).

The next three sections address these considerations and make recommendations on dimensions and methodologies for ALM training evaluations that should get more attention in the future.

4.4.1. Broadening evaluation scope

The focus of evaluation should be broadened and should ⁽¹⁰²⁾:

- (a) address the determinants of various outcomes, compared across subgroups of the labour market and explore participants' motivational factors;
- (b) be carried out over longer periods than now is the case, to see if the short-term effects on earnings and employment

⁽¹⁰¹⁾ One of the fundamental evaluation problems is that the more time passes, the more difficult it is to follow a sample and to attribute observed effects to the intervention.

⁽¹⁰²⁾ Based on the following literature: Fay (1996, cited in Walsh and Parsons, 2004), Martin and Grubb (2001: 16), Walsh and Parsons (2004), Hujer et al. (2004a), Grubb and Ryan (1999).

- prevail in the long-run and if there are some delayed effects ⁽¹⁰³⁾;
- (c) prefer increased employment probability or reduced unemployment duration as outcome variables. Earnings and wages might be questionable outcome variables in the European context due to the fact that welfare state and minimum wage regulations result in distortions between employment status and earnings;
 - (d) extend the scope of effects (or side-effects) for consideration to non-material ones (increased self-worth, better health, wider social and behavioural gains, further learning experiences, etc.);
 - (e) be foreseen from the outset when an intervention is designed so that the requisite information is gathered in a continuous process;
 - (f) attempt to assess better the structure, content and design of training courses as well as whether they can be adapted to different circumstances, such as changing labour-market needs.

4.4.2. Opening the black box: the question of why?

While evaluations tell us a lot about what works, or not, they are not so investigative of another equally important and related question: why? Why do certain programmes work for some groups and not for others? Why do some circumstances increase the likelihood of a programme working? Why do specific programme design and characteristics work well, work better or not work? Finding answers to these questions is a central issue in designing cost-effective public training programmes. Formative evaluation and more qualitative methods are required to address these questions because they deepen understanding of the way an intervention operates and participants react to it (see Part 2, Section 1.4.2).

An important step is to establish what happens in implementation. Before evaluating outcomes, evaluations should consider whether a programme has been established as intended. Is the programme coherent with its initial design and intentions or have they been changed? Was the programme somehow adapted to local circumstances? Possible methods of investigating these issues are inspections of facilities and activities, questionnaires and interviews with providers, trainers and trainees (Grubb and Ryan, 1999: 70; see also Section 1.2 in Part 2).

Another important aspect is to consider the learning process (which is central to any training programme). Nicaise and Bollens (1998, cited in Descy and Tessaring, 2001a: 364) note that compared to the substantial body of work on the evaluation of training for the unemployed, '[...] theoretical or empirical work concerning the relative contribution of programmes' organisational, curricular and instructional characteristics and the interdependencies between these characteristics is much less developed.' Evaluation often fails to consider the teaching and learning process. It tends to simply assume that the learning process happened and resulted in the skills the programme intended to develop. In addition, few studies are concerned with the most effective teaching environments and methods ⁽¹⁰⁴⁾, failing to consider that it is an important aspect of programme design. One cannot conclude on the (non-) effectiveness of a certain type of training without establishing whether any learning, let alone quality learning took place.

According to Grubb and Ryan (1999: 71), to judge the quality of the learning process and teaching methods, the preferred research approach should be a qualitative investigation of the motives for and experiences of training. This is typically done through case studies of particular training programmes, intensive

⁽¹⁰³⁾ This aspect is crucial because programmes have been established to address systemic and enduring problems such as unemployment, social exclusion, transition from school to work, etc. These are the kinds of problems that require long-term solutions. It must, therefore, be established if the short-term results of a programme (what is generally captured by evaluation) are sustainable.

⁽¹⁰⁴⁾ Limiting itself to comparing on-the-job and off-the-job training. E.g. pedagogic and didactics are not considered.

interviews with participants, providers and administrators. By using these methods ⁽¹⁰⁵⁾, reasons for failure and for success can be explored and the results used to improve current or future programmes.

Exploring programme design and implementation can also help to determine if the causes of a programme's (non-) effectiveness are endogenous (e.g. appropriateness and quality of the learning process) or exogenous (e.g. the existence of jobs corresponding to the skills developed in the programme). These dimensions should complement those evaluating the effectiveness or the cost-benefits of a programme ⁽¹⁰⁶⁾.

Evaluations that examine only the final outcomes of programmes in economic terms, such as earnings or employment, or even in non-economic terms, such as better health or reduction in crime, assume that this is all that is necessary to know in order to decide whether a programme should be maintained, expanded or terminated. Improving the quality of a programme in order to enhance its positive outcomes should also be considered as a key evaluation goal but it requires more extensive and in-depth information on the programme. It requires that one understands why it works rather than only whether it works. This formative (and more qualitative) kind of evaluation, in combination with the summative (and more quantitative one), contributes to obtaining a fuller picture and therefore to improving the effectiveness and the quality of interventions.

4.4.3. Towards a systemic approach for evaluating ALMPs

The type of evaluation that has been discussed in this chapter is conducted in what can be called a 'programme' perspective. ALM training is designed and implemented within a defined period and then evaluated, asking

whether it benefited the participants or had an impact on unemployment.

These evaluations, because of the methods they often employ and their focus on assessing the effectiveness of single projects and programmes, reinforce the tendency to design policies that are limited in scope and target population, rather than aiming at long-term institution building. Those former policies, although they can be relatively successful, cannot do much for a country's growth and development. They gear policy-makers towards dropping old programmes and developing new ones.

Grubb and Ryan (1999: 109) comment: '[...] a programme is conceived of as something that can be created relatively quickly, introduced among the other institutions of a society, evaluated as a discrete entity, expanded or contracted. We call this a "project" or "programme view", because of its tendency to think of a VET programme as self-contained and independent. Then, what is conventionally called "programme evaluation" assesses the effect of the programme only, independently of any surrounding policies and institutions. We contrast this with what we call a "systems" view.' They elaborate later on (ibid.: 116): '[...] only rarely do the evaluation of VET programmes conceive of a larger system of programmes [...] a system that might develop slowly over time. [...] A corollary is that more energy is put into developing new programmes and evaluating them – and then abandoning them or trying new approaches – rather than continuing to develop institutions over longer periods of time [...] This suggests that the entire evaluation enterprise, in its "programme" focus, is part of an incoherent and fragmented approach that is unlikely to lead to more effective VET policies over time. A system's perspective would, on the other hand, encourage thinking not about individual projects, but about widely available

⁽¹⁰⁵⁾ For a presentation of various qualitative methods, please see Section 1.2 in Part 2.

⁽¹⁰⁶⁾ A TSER transnational research project dealing with the identification of process variables influencing the outputs and the outcomes of training for the long-term unemployed is presented in Descy and Tessaring (2001a: 364-365) and Brandsma (1999). It concludes that the closer training is to working life and the later in the course of the programme job search training is provided, the more chances of finding a job increase. The effect of flexible curricula and of guidance and counselling were not clear.

programmes that are linked to one another and institutionalised.’

The basic reason for adopting a systemic approach in policy design and evaluation is that any policy will interact with the rest of the system (institution, organisations and actors) and with other systems (such as production). In addition various types of programmes, led in parallel, also interact. A system perspective emphasises the need for more coherence between the various VET interventions, ALMPs as well as other social policies.

Issues of internal consistency (is this programme adding up, conflicting or cancelling other active or passive measures?) and of external consistency (is this programme consistent with the institutions in place, e.g. does it take into account signalling patterns in the labour market?) are to be considered in both programme design and evaluation ⁽¹⁰⁷⁾. Close to these issues is the one of transparency, i.e. the ability of workers and employers to understand the role of a specific programme in the vast range of learning opportunities offered in education, VET and various ALMPs.

Considering whether a successful programme could be expanded in terms of target groups, or geographical terms also requires a system perspective, i.e. are the conditions that lead to success replicable? Would potential benefits also apply considering the increased costs of an expanded programme? Would it induce saturation in the labour market for ex-participants? Would the expanded programme be consistent with the other practices, values and institutions in the country?

Finally, in a system perspective it is worth asking whether a programme is linked to other learning opportunities. ALM training has a generally narrow skills focus. To allow individuals to achieve both the levels and the varieties of competences necessary throughout their lives, one should make sure that each programme can be potentially linked to other programmes ⁽¹⁰⁸⁾ (Grubb and Ryan, 1999: 118).

Policy-makers and evaluators should be expanding their views regarding the nature of useful interventions. The development of programmes and measures has to be seen in a context where they relate one to another, in addition to their individual effectiveness. This should also be reflected in evaluation designs.

In conclusion, systemic evaluating means considering whether a programme is successful because of its ability to be replicated, articulated with other VET programmes, connected to employers’ hiring practices, and otherwise related to other established practices and institutions. This approach could lead to the conclusion that even a programme with positive and significant outcomes in a ‘programme view’ may be ineffective in a ‘system perspective’.

Systemic evaluation is still in its infancy, but, as discussed in Chapter 2, has been applied to the evaluation of VET (especially in central and eastern Europe countries). The methodology that is developed there could be transferred into the field of evaluation of ALMP, particularly training.

Grubb and Ryan (1999: 119) conclude: ‘Evaluation is intended to reflect the effectiveness of a particular programme independently of other social practices and institutions that surround it. The basis evaluation question is: what is the difference – for individuals, for employers, for other workers, for an economy as a whole – due to having a particular programme in place, assuming that other programmes and policies stay the same? While it is certainly an important question, it may generate misleading answers if other policies or institutions are necessary for a VET programme to be successful.’

⁽¹⁰⁷⁾ See Section 2.3 for a more in-depth discussion of internal and external consistency.

⁽¹⁰⁸⁾ Recognising and validating the learning results of ALM training could constitute a step in this direction.

4.5. Conclusion: continuous improvement of employment measures and policies

ALMPs are being used more frequently in all EU countries, training being one of their key measures. Demonstrating their efficiency and improving their quality should be a concern for policy-makers.

The traditional evaluations of ALMPs 'compare' explicit programme goals with measurable outcomes (mainly in terms of employment probability and earning) and look at cost-benefits. They allow estimation of the relative effectiveness and/or efficiency of different interventions. Most of them are quasi-experimental, with emphasis on the econometric elaboration of programme outcomes. Because traditional evaluations do not open the black box, their results are generally limited in terms of indicating possibilities for change and improvement of programme design and implementation. Instead, they lead to a decision either to stop or to continue programmes. They concentrate on what works rather than trying to answer why it works or not. They also tend to neglect the interaction between various policy interventions and their cumulative impact (Schmid et al., 1996: 2).

Evidence-led public policy is more in line with a holistic kind of evaluation, combining formative and summative approaches, thus collecting the breadth and scope of information needed to provide feedback on quality. 'Some of the best studies are those that take a wider methodological perspective, certainly using [quasi-] experimental approaches where feasible, but complementing this with the use of administrative data and more qualitative information on for example, processes and the perceptions of programme participants. It suggest that evaluation hitherto has been more "academically" driven rather than "policy" driven, which is not so much a criticism of the former as a lack of proper attention in the latter.' (Walsh and Parsons, 2004, Section 5.2).

Grubb and Ryan (1999: 32) also think that although measurements, using (quasi-) experimental methods, of employment-related outcomes are probably the most important results of VET programmes and 'deserve priority over fluffier measures' (such as satisfaction of participants), multiple approaches are desirable. This is particularly true for improving the quality of programmes and enhancing the public debate around VET.

Evidence-led policy-making requires information on the circumstances, both exogenous and endogenous, that lead to programme quality, positive outcomes and cost-effectiveness. These requirements go beyond what common evaluation practices offer. In the future, a systemic approach to evaluation should be developed alongside existing techniques; this should broaden the current perception of effectiveness by focusing also on the way programmes and measures interact with other interventions and existing institutions, can be expanded, replicated and can lead to further learning opportunities. Policy evaluation should define the criteria and provide supporting empirical evidence on which policy types and which policy mix promise superior solutions to a society's problem (Schmid et al., 1996: 12).

Finally, evaluation should also be used to trigger public debates. In this respect, one must admit that many evaluations of ALMPs are not easily understandable for the non-scientific community because of the complexity of their methods and interpretation of their results. According to Grubb and Ryan (1999: 29), one way to stimulate public debate around the effectiveness of ALM training is to involve interest groups in formulating evaluations. They can contribute towards specifying what stages of a programme are to be evaluated, what outcome measures are to be emphasised and what issues of quality are to be investigated. This would contribute to building a sense of ownership of evaluation, results promoting their use.

PART 4

Impact and benefits of education and training

Abstract

Education, training, and skills – or human capital – are seen as key determinants for economic success and for achieving individual and social goals. At micro level, employability and quality of life are expected to improve for people with higher levels of skills and competences. Education training, skills and competences are supposed to be major factors in the economic performance of companies. At macro level, investments in human capital are expected to have considerable benefits in terms of enhanced economic growth and social cohesion.

These assumptions are increasingly being put to the test. Education and training are assessed in terms of their costs and benefits, individual employment prospects and careers, productivity and economic performance of companies and contribution to economic growth, employment and social cohesion. This arises from increasing cost pressure and competition for enterprises, persistent tight public budgets and the expectations of individuals that the high outlay in time, effort and costs for education and training should pay off in terms of career, earnings and work status.

However, the analysis and empirical verification of causal links between education and training and various benefits at all levels of society is complex for various reasons.

First, different ‘agents’ – individuals, enterprises, regions, the State – who invest in, and benefit from, education, training and human capital have to be distinguished. Second, benefits ‘spill-over’ to agents other than those investing in or undergoing education and training. Third, the nature of benefits is heterogeneous. They can be economic or social, material or non-material, monetary and non-monetary, direct and indirect, positive or negative.

Finally, economic models or advanced econometric techniques will never be able to grasp in a single framework all links, relationships and behaviours of different actors, let alone more societal and value-laden phenomena such as citizenship, social cohesion and trust. They cannot encapsulate all kinds of impacts but can just try to identify specific causes and effects. Equally, it is difficult to include more qualitative, i.e. not quantifiable or even unobservable, characteristics and effects in such models.

In empirical analysis, the usual approach is to isolate the triggering event (e.g. investments in education and training) from other causes which might also have an impact on the outcome we are interested in, for example increase in earnings, reduction of unemployment or social inclusion of specific groups. This can be done by eliminating other factors of influence: in macroeconomics by applying the *ceteris paribus* clause which holds ‘all other things constant’; and in microeconomics by ‘controlling for’ the effects of other intervening variables. However, we should always be aware of the limitations and constraints on how this information has been gathered and by which methods it has been analysed.

This part will discuss recent research on the diverse benefits of education, training and skills – of human capital. When considering the structure of this part there were several options. The option for which we decided was to place the recipients of benefits in the foreground: society as a whole, companies and individuals. However, overlaps between chapters of this part could not be avoided as most studies address benefits across several recipients and levels.

Chapter 1 addresses the contribution of human and social capital to economic growth at macro and regional level. It focuses particularly on the new theories of economic growth which see human capital investments, research and development (R&D) activities and the dissemination of knowledge as the main determinants of long-term economic growth. More recent research extends this perspective by reasoning about the quality of schooling and addressing social capital and social infrastructure as decisive framework conditions for profitable investments in human and physical capital.

Chapter 2 discusses, at macrosocial level, the non-material and wider benefits of education and training such as crime reduction, trust and social cohesion. Beyond their social impacts, these benefits are also expected to foster economic growth and performance. However, most of these macrosocial benefits are not directly, but indirectly linked with education and training.

Chapter 3 presents evidence on the benefits of education, training and skills for companies. In contrast to traditional thinking, that companies only invest in specific training which is not 'marketable' but only of benefit for the training firm, there is increasing evidence that companies also invest in more general training. This is particularly the case in modern, 'high performance work systems' where training is an ingredient part of human resource management and is expected to contribute to a company's productivity, profitability and competitiveness.

Chapter 4 addresses material and non-material benefits for the individual. In addition to common approaches to calculating the economic rates of return (RoR) on education and training, this chapter discusses educational benefits – both material and non-material – in life-course and biographical perspectives, which are relatively new strands of education and training research.

Table of contents

1. The determinants of economic growth at macro and regional level	196
1.1. Theories of economic growth	196
1.2. Human capital and economic growth: empirical evidence	198
1.2.1. Results of neoclassical growth accounting studies	198
1.2.2. Evidence from endogenous growth studies	199
1.2.2.1. Levels of education and training: effects of marginal investment	204
1.2.2.2. The contribution of vocational training to growth	206
1.2.2.3. Quality of schooling	207
1.2.3. Education, training and economic performance at regional and local level	208
1.2.3.1. Inter-regional inequalities	209
1.2.3.2. Regional development and growth: empirical evidence	209
1.3. Social capital and economic performance	210
1.4. Discussion	212
2. Non-material benefits and externalities of education and training	214
2.1. Features of macrosocial benefits	214
2.1.1. Attribution of social benefits	214
2.1.1.1. Non-identifiability	214
2.1.1.2. Relational or positional nature	215
2.1.1.3. Benefits at individual and at system level	215
2.1.2. Non-material benefits and externalities	216
2.1.3. Education, training and crime	217
2.1.4. Education, training and social cohesion	218
2.1.5. Social capital and trust	219
2.2. Evaluating macrosocial benefits: an empirical assessment	221
2.2.1. Social cohesion and macrosocial indicators	222
2.2.2. Assessment of macrosocial benefits using microsocial data	224
2.3. Discussion	225
3. The significance of education and training for company performance	227
3.1. General or specific training: who pays, who benefits?	227
3.2. Timing of training and impacts	229
3.3. Skills and company performance	230
3.3.1. Impact of education and training on productivity: review of research findings	230
3.3.2. Small and medium enterprises	236
3.3.3. Human resource management and high performance work systems	239
3.3.3.1. HRM and in-company training	239
3.3.3.2. Training, technological change and innovation	243
3.3.4. The market value of a firm	244
3.3.4.1. Indicators for the market value	244
3.3.4.2. Training investments as information on market value	245
3.4. Discussion	247

4. Individual benefits of education and training	247
4.1. Monetary returns on investment in human capital	248
4.1.1. Private returns on education	248
4.1.2. Private vs social RoR	251
4.2. Benefits of education and training in the individual life course	252
4.2.1. Research findings at national and European level	255
4.2.1.1. Monetary returns and lifetime income	256
4.2.1.2. The effects of education and training on labour-market participation	258
4.2.1.3. Education, training and transitions	259
4.2.1.4. Educational benefits in a cohort perspective	261
4.2.1.5. Educational benefits, social background and gender	262
4.2.1.6. Non-material benefits	264
4.2.1.7. Biographical research	265
4.3. Discussion	266

List of tables, figures and boxes

Tables

4.1.	Accounting economic growth and level: results of selected studies	199
4.2.	Impact of human capital on economic growth: selected cross-country growth regressions	200
4.3.	The level of interpersonal trust in OECD countries, 1990-91/1995-96	221
4.4.	Correlation of education and income distribution with social cohesion at aggregate level	223
4.5.	The impact of education and training on company performance – selected studies	232
4.6.	Provision and costs of training by size of enterprise, 1999	237
4.7.	Selected studies on training in SMEs	238
4.8.	Studies on human resource management and high performing work systems	242
4.9.	Marginal RoR to education in Europe, 1995	250
4.10.	Private and social RoR on education in selected OECD countries by gender 1999-2000 (in percentage points)	251
4.11.	Selected studies on benefits of education and training from a life course perspective	254

Figures

4.1.	The contribution of human capital enhancement to labour productivity growth 1990-2000 (average annual percentage change)	202
4.2.	Effects of education on social cohesion in four countries	225

Boxes

4.1.	Impact of the quantity and quality of schooling on growth	208
4.2.	Social capital and economic payoff – results from the world values survey	211
4.3.	Knowledge, intellectual capital and intangible assets	211
4.4.	Examples for non-material benefits and human capital externalities	214
4.5.	Education and training: public or private goods?	215
4.6.	Health and environmental externalities	216
4.7.	The concept of social cohesion	218
4.8.	Etgace – Education and training for active citizenship in Europe	219
4.9.	The World values survey	220
4.10.	Education and social cohesion: indicators, countries and methods of analysis	222
4.11.	Data sources, countries and indicators for the analysis of social benefits using microdata	224
4.12.	Surveys carried out by the ASTD	231
4.13.	High performance work systems (HPWS)	240
4.14.	The Cranet network	241
4.15.	Public funding and private returns on education (PURE)	249
4.16.	Life courses in the globalisation process (Globalife)	260
4.17.	Socioeconomic determinants of healthy ageing (SEdHA)	265

1. The determinants of economic growth at macro and regional level

The 2000 Lisbon European Council ⁽¹⁾ set the goal to make the EU, by 2010, ‘the most competitive and dynamic knowledge-based economy in the world, capable of sustainable economic growth with more and better jobs and greater social cohesion’. This goal, made concrete in subsequent summits, accompanied by activities of the European Commission and of the Member States and by agreements of the European Social Partners, put the emphasis on investments in, and development of, lifelong learning, education and training. This reasoning relies on the assumption that investing in, and making the best use of, human resources leads to economic growth and social cohesion ⁽²⁾.

Against this background, this chapter reviews research in the macroeconomic benefits of human capital. It discusses in particular research findings concerning the contribution of education, training and skills to economic growth at international, national and regional level. It includes a discussion on the role of social capital which is assumed also to have clear impacts on economic growth.

However, as the review of research will show, it is easier to ask questions than to provide answers. The range of findings concerning the ‘sources of growth’ is extensive, depending on the country considered, the data available, the period observed, the methodology used and the theories on which growth models are based. Furthermore, most research only provides crude answers to the question of which types of skill, knowledge and competences and which level of education and training have the highest ‘predictive’ capacity for economic growth. Nevertheless, and thanks to major

advancements in theory, methods and data sources within the past few years, recent research is unambiguous on the considerable benefits of education and training investments, and rising skill levels of populations, on economic growth.

1.1. Theories of economic growth

The contribution of education/training or human capital ⁽³⁾ to economic growth is the domain of macroeconomic research which attempts to investigate the ‘sources of growth’ by taking into account all production factors (physical capital, labour, land) that influence the growth of GDP (or GNP; see Box 2.4 in Part 2) over a period of time. Most of this research uses models based on economic growth theories. In this context we should remember the important distinction between studies based on neoclassical growth theories and those based on the new theories of endogenous growth ⁽⁴⁾.

Traditional neoclassical growth models look at the quantity of (physical and human) capital. Although augmented models in the neoclassical tradition also include the quality of labour, i.e. human capital, they do not seek to explain productivity and technological change but are treated as exogenous (residual) factors. These models assume that a one-off increase in the human capital stock will lead to a one-off increase in the economic growth, until productivity per working hour has reached its new (and permanently higher) ‘steady-state’ level (Sianesi and van Reenen, 2002: 40).

In contrast, new growth models include the quality of production factors such as knowledge, human capital and R&D and seek to explain productivity and technological

⁽¹⁾ For more information on the Lisbon strategy and the various follow-up activities see: http://europa.eu.int/comm/lisbon_strategy/index_en.html [cited 19.7.2004].

⁽²⁾ Note that the DG Education and Culture of the European Commission launched, in summer 2003, several projects in economics of education, dealing with financing of higher education, returns on various types of investment in education and training; and the creation of a European expert network.

⁽³⁾ See for definitions Part 2, Box 2.5.

⁽⁴⁾ See for a more detailed discussion Part 2, Section 2.1.1.

progress explicitly in their models; they are considered 'endogenous' factors.

'New theories of endogenous growth [...] argue that the same one-off increase in human capital will be associated with a permanent increase in the growth rate. The social benefits of education will clearly tend to be much greater in this case.' (ibid.) Endogenous growth models set human capital and R&D activities at the centre of their framework; R&D being closely associated with human capital (research workers) and knowledge.

Theories predict that productivity (GDP per capita) is determined by the resources – education and training/human capital and physical capital – devoted to R&D and the diffusion of R&D to other companies and countries. The results of successful R&D (for example a product or process innovation) or the publication of a patent specification, disclose information that may be of use to competitors or to companies in other product areas which use similar technologies. This knowledge, therefore, cannot be kept secret but will spread out to other companies and across countries. Many studies that have investigated the role of knowledge spill-over have found that they appear to be statistically more important than the firm's own R&D or patents. Section 1.2.2. will present some results of recent research in this field.

We should note that economic growth models view education and (sometimes) training as investments in human capital. Past investments in education and training and the resulting stock or level of human capital are most often proxied by the number of years of schooling or the levels of educational attainment of the workforce. Thus, enrolments in education and training are seen as flows which lead to increased educational attainment and the accumulation of the human capital stock. However, most empirical growth models adopt a rather narrow interpretation of human capital. Most empirical studies focus on education which, in most data sets used (e.g. based on the ISCED classification), also

includes formal initial and partly continuing vocational training (CVT). Some studies which particularly investigate the effects of training delimit training to on-the-job or off-the-job training, which in most cases have an informal or non-formal character. Only few attempts have been made to incorporate a wider notion of human capital which includes knowledge, skills, competences and similar attributes acquired by formal and non-formal learning ⁽⁵⁾.

Economic growth models are often criticised for their non-spatial design, which assumes all production factors to be mobile across regions. The assumptions of a frictionless diffusion of technological knowledge and of unlimited mobility of workers are particular topics of debate. Results of spatial studies, however, indicate considerable disparities between regions. This particularly concerns unequal development of skills and human capital which most likely is pronounced spatially among regions, as well as among nations. This implication is in stark contrast to neoclassical models of economic growth which predict convergence across regions and countries. The basic assumption of these models is one of perfect markets, which immediately balance out disequilibria and imply unlimited mobility of all production factors, including labour ⁽⁶⁾.

In contrast to the assumptions in neoclassical models, research findings display no convergence but considerable cross-country and cross-regional differences in terms of economic growth and human capital of populations. What makes some countries or regions accumulate more human capital than others or what makes them more efficient than others in the use of such inputs? Given the limitations of macroeconomic growth models in terms of defining human capital, an increasing body of research is devoted to exploring the significance of social capital or social infrastructure for economic performance and to investigating the links between human capital, social capital and economic prosperity. Social capital and infrastructure can be seen

⁽⁵⁾ See also Descy and Tessaring (2001a, in particular Part 1, Chapter 4 and Part 3, Chapter 2).

⁽⁶⁾ Section 1.2.3 discusses some aspects of growth research at regional and local levels.

as both independent framework conditions for economic success and outcomes of increased education and training investments. This view associates higher levels of human and social capital, and of social infrastructure, with better environment, higher levels of health, trust and greater social cohesion. All these effects are expected to feed back into faster economic growth. Some of these issues will be discussed in Section 2.3. A more detailed discussion on these and other macrosocial benefits is to be found in Section 3.4.

1.2. Human capital and economic growth: empirical evidence ⁽⁷⁾

Over the last three decades considerable research work has been devoted to examining sources of economic growth and, in particular, the contribution of human capital to the level and growth of the gross domestic product (GDP) or GDP per capita ⁽⁸⁾. Much of the work deals with different growth model specifications.

The first generation of cross-country studies based on endogenous growth theories ⁽⁹⁾ indicated that R&D and the accumulation of human capital by investments in education and training play a central role in determining economic growth (de la Fuente and Ciccone, 2002: 32). However, some later studies questioned the link between education and growth ⁽¹⁰⁾. More recent research argues, however, that these results were largely due to

poor data and various econometric problems. Studies that use improved data sets ⁽¹¹⁾ suggest that investments in education/training in general have a substantial impact on growth, though not always nor everywhere ⁽¹²⁾.

Recent empirical work seeks to go beyond cross-sectional data for a large number of countries. For smaller groups of countries, such as those in the OECD area where the quality of educational data is perceived to be both higher and more reliable, some researchers have attempted time series analyses. Other (few) studies have combined cross-sectional data with time series to produce panel data which allow analysis of country-specific effects over time (for a review see Wilson and Briscoe, 2004).

1.2.1. Results of neoclassical growth accounting studies

Table 4.1, mainly drawn from Sianesi and van Reenen (2002) ⁽¹³⁾, compiles results of some growth accounting studies carried out in the past decade to calculate the contribution of human and physical capital to economic growth ⁽¹⁴⁾. They assess the contribution of inputs to the output, either economic level or economic growth. The input factors are physical capital and human capital and are set against the total factor productivity which denotes the efficiency with which these inputs are used. However, total factor productivity is a residual and remains unexplained in neoclassical accounting models.

Studies in the neoclassical tradition, in particular those using an augmented form (see

⁽⁷⁾ This section is mainly based on the following summarising studies: the contribution by Wilson and Briscoe (2004) for the third research report; Sianesi and van Reenen (2002) and de la Fuente and Ciccone (2002). Further literature references can be found in these articles.

⁽⁸⁾ For definitions see Part 2, Box 2.4.

⁽⁹⁾ See among others Landau (1983), Baumol et al. (1989), Barro (1991) and Mankiw et al. (1992).

⁽¹⁰⁾ Studies that report largely negative findings include Kyriacou (1991), Benhabib and Spiegel (1994), Pritchett (1999, whose first version is from 1995), Islam (1995) and Caselli et al. (1996).

⁽¹¹⁾ According to Wilson and Briscoe (2004), most studies use the Penn World Table series produced by Summers and Heston (1991). This data set has been updated and extended by Barro and Lee (1996).

⁽¹²⁾ See for instance de la Fuente and Doménech (2000), Krueger and Lindahl (2001), Cohen and Soto (2001) and Bassanini and Scarpetta (2001).

⁽¹³⁾ Their paper can be downloaded under <http://www.ifs.org.uk/workingpapers/wp0205.pdf>. The paper also includes a detailed review of the main studies referred to below.

⁽¹⁴⁾ Note for the interpretation of the tables that follow the distinction between 'percent(age)' and 'percentage points': percentages indicate the relative change of absolute figures (e.g. GDP, number of people) within two periods whereas percentage points denote the difference between two percentages. An example: the difference of the GDP growth rates 1995/94 (3.5 percent) and 1996/95 (4.5 percent) is 1.0 percentage point.

Table 4.1. Accounting economic growth and level: results of selected studies

Study	Output	Total factor productivity	Input	
			Physical capital	Human capital
Jorgenson and Fraumeni (1992)	Growth accounting: output growth rates, US 1948-86	17 %	22 %	61 % (labour input) of which 42 % is accounted by labour quality. Therefore labour quality accounts for 26 % of economic growth
Mankiw et al. (1992)	Level accounting: cross-country differences in output per worker, 98 countries in 1985	22 %	29 %	49 %
Klenow and Rodriguez-Clare (1997)	Level accounting: cross-country differences in output per worker, 98 countries in 1985	67 %	29 %	4 %
	Growth accounting: cross-country differences in 1960-85 growth in output per worker, 98 countries	85-90 %	3 %	6-12 %
Hall and Jones (1999)	Level accounting: cross-country differences in output per worker, 127 countries in 1988	61 %	17 %	22 % (educational attainment for the population over 25)

Source: Sianesi and van Reenen (2002: 19 f.).

previous section), point to the importance of investments in physical and human capital (Jorgenson and Fraumeni, 1992; Mankiw et al., 1992). An increase in labour input accounts for 61 % of US growth; less than half of the economic growth is attributable to increases in human capital, such as level of educational attainment. More recent studies have questioned these results and their underlying methodologies, claiming that not simple labour input but the productivity of all production factors – which remains unexplained (residual) – is by far (more than 60 %) the most important determinant of growth.

Sianesi and van Reenen (2002: 17 f.) argue that ‘the “human capital” aspect of the labour input used is often a combined measure of various educational, demographic and labour force variables (i.e. account is taken of changes in the age, sex and educational composition of the workforce, as well as of hours of work). Finally, [...] “accounting is no explanation”. Apportioning income or income growth to measured and

unmeasured “inputs” provides no insight as to the mechanisms which may underlie such contributions. And as to those studies showing the overwhelming importance of total factor productivity, results need to be further explored as to examine the fundamental sources of such a factor.’

1.2.2. Evidence from endogenous growth studies

Cross-country growth regressions that are based on endogenous growth theories intend to explain the sources of growth by seeking to assess – in addition to the contribution of physical capital – the impact of human capital and of R&D, on productivity and a country’s economic growth (Sianesi and van Reenen, 2002: 21). Human capital is mostly proxied by educational levels, years of schooling or school enrolment rates.

Many of the various studies using cross-country growth regression cannot be compared directly because of different dependent variables ⁽¹⁵⁾, different human

⁽¹⁵⁾ For example: real per capita GDP growth rates; overall real GDP growth rates; growth of labour productivity or of total factor productivity; logarithm of the ratio of real (per capita or overall) GDP in two or more periods.

Table 4.2. **Impact of human capital on economic growth: selected cross-country growth regressions**

Study	Dependent variable	Human capital proxy	Flow/stock
Barro (1991)	Annual growth rate of real GDP per capita, 1960-85	School enrolment rate in primary and secondary education, in 1960	flow
Levine and Renelt (1992)	Annual growth rate of real GDP per capita, 1960-89	Secondary school enrolment rate in 1960	flow
Murphy et al. (1991)	Growth rate of real GDP per capita, between 1970-85	Primary school enrolment rate in 1960	flow
Barro (1997)	Growth rate of real GDP per capita, periods: 1965-75, 1975-85, 1985-90	Average years of attainment for males aged 25 and over in secondary and higher schools at the start of each period	initial stocks in 1965, 1975 and 1985
Hanushek and Kim (1995)	Growth rate of real GDP per capita, between 1960-90	Average years of secondary schooling of adult male population at beginning of period	initial stock
Gemmel (1996)	Annual growth rate of real GDP per capita, 1960-85	Constructed human capital stock in 1960 and human capital annual average growth rates at primary, secondary and tertiary levels	initial stock and annual flows
Judson (1998)	Growth rate of real GDP, five-year averages, 1960-90	Growth of constructed measure of human capital stock	period flows
Englander and Gurney (1994)	Growth of labour productivity (and total factor productivity, over four time periods (OECD))	Enrolment rates in secondary schools	flow
Barro and Lee (1994)	Growth of GDP per worker	Average years of secondary schooling of adult male population at beginning of period	initial stock
Benhabib and Spiegel (1994)	Growth of GDP per capita	Human capital stock estimates: average level of log human capital over the period (log of average level of human capital; log of average levels)	average stock
Graff (1996)	Annual GDP per worker, 1965-90	Average number of workers' years of schooling at lower, intermediate and higher levels	initial stock

Illustration of impact

A 1 percentage point increase in primary (secondary) school enrolment rates is associated with a 2.5 (3.0) percentage points increase in per capita GDP growth rate.
A 1 percentage point increase in secondary school enrolment rate is associated with a between 2.5 and 3.7 percentage points increase in per capita GDP growth rate.
A 1 percentage point increase in primary school enrolment rate is associated with a 2.2 percentage points increase in per capita GDP growth rate; sub-sample for OECD countries: not significant.
An extra year of male upper-level schooling is associated with a 1.2 percentage point increase in per capita GDP growth rate.
An extra year of male secondary schooling is associated with a 0.36 percentage point increase in per capita GDP growth rate.
For OECD sub-sample: a 1 percent increase in tertiary human capital stock is associated with a 1.1 percentage point increase in per capita GDP growth rate. A 1 percentage point increase in tertiary human capital growth is associated with a 5.9 percentage points increase in per capita GDP growth rate. For the poorest group of developing countries primary human capital is most significant whereas for the intermediate group of developing countries it is secondary human capital.
A 1 percentage point increase in human capital growth is associated with an 11 percentage points increase in GDP growth rate.
A 1 percentage point increase in secondary school enrolment rate is associated with an around 1.5 percentage point increase of productivity growth.
An additional year of male secondary schooling is associated with a 1.4 percent increase in per worker GDP growth.
A 1 percent increase in the stock of human capital is associated with a 12 to 17 percent increase in per capita GDP growth (but not always significant)
A 1 percent increase in human capital stock at lower/intermediate/higher level is associated with a 1.5/4.2/4.2 percent increase of GDP per worker. For countries with lower educational imbalance ⁽⁶⁾ the coefficients for the three levels are: 1.5/5.0/6.9 percent; for countries with higher educational imbalance the respective coefficients are: 1.5/0.5/-2.1 percent.

capital measures ⁽¹⁶⁾ and different country samples. Furthermore, as Sianesi and van Reenen (2002: 22) point out, ‘it is important to understand that the main aim of such studies is to identify statistically significant and possibly robust relationships between various factors and economic growth. [...] The main message the authors seek to convey to the reader is that a given factor does indeed positively – or negatively – affect growth, and is more – or less – important than another. Methodological and especially data constraints seem to severely hinder a precise numerical quantification of the effects, so that the actual magnitude of the estimated effect is almost invariably ignored.’

Appropriate data and long time series are required to calculate the components of economic growth irrespective of short- or medium-term cyclical or non-linear developments caused by business cycles or ‘shocks’ such as wars or energy crises). Many available data only indicate tendencies and only marginally provide reliable answers at macro level. Furthermore, analyses for a particular country should control for (eliminate) endogeneity and other biases ⁽¹⁷⁾ (Part 2, Section 2.1.1.3).

Table 4.2 summarises selected studies on cross-country growth regressions.

The reviews by Wilson and Briscoe (2004) and Sianesi and van Reenen (2002) of growth research suggest in general that investments in, and level of, human capital – together with investments in physical capital and R&D – play a prominent role in explaining the economic growth of a country as well as differences between countries. However, there remains

⁽¹⁶⁾ Concerning the stock (e.g. average years of education or educational attainment) or flows of human capital (e.g. enrolment rates). Some studies apply own measures of human capital stock and accumulation which are constructed to overcome some shortcomings of commonly used proxies. These studies, however, have the disadvantage of not being comparable with other countries or other studies and thus lack immediate policy interpretation.

⁽¹⁷⁾ An endogeneity bias occurs if it is not education and training investment which contributes to a country’s growth of domestic product, but if economic growth and performance lead to further investment in education and training (reverse causality).

Table 4.2. (cont'd)

Study	Dependent variable	Human capital proxy	Flow/stock
Mankiw et al. (1992)	Growth of GDP per working-age person	Average percentage of working-age population in secondary school, 1960-85	period flow
de la Fuente and Doménech (2000)	Annual GDP per worker, 1960-90 (OECD)	Average number of years of schooling of the adult population	stock
Bassanini and Scarpetta (2001)	Annual GDP per working-age person, 1971-98 (OECD)	Average number of years of schooling of the adult population	stock
Barro (2001)	Real GDP per capita, 1965-75, 1975-85, 1985-95	Years of schooling at primary, secondary and tertiary levels	stock

(*) Indicator for educational imbalance: ratio of enrolment in tertiary education / enrolment in primary education.
 Source: Sianesi and van Reenen (2002: 23 ff.) and updates.

Figure 4.1. The contribution of human capital enhancement to labour productivity growth 1990-2000 (average annual percentage change)

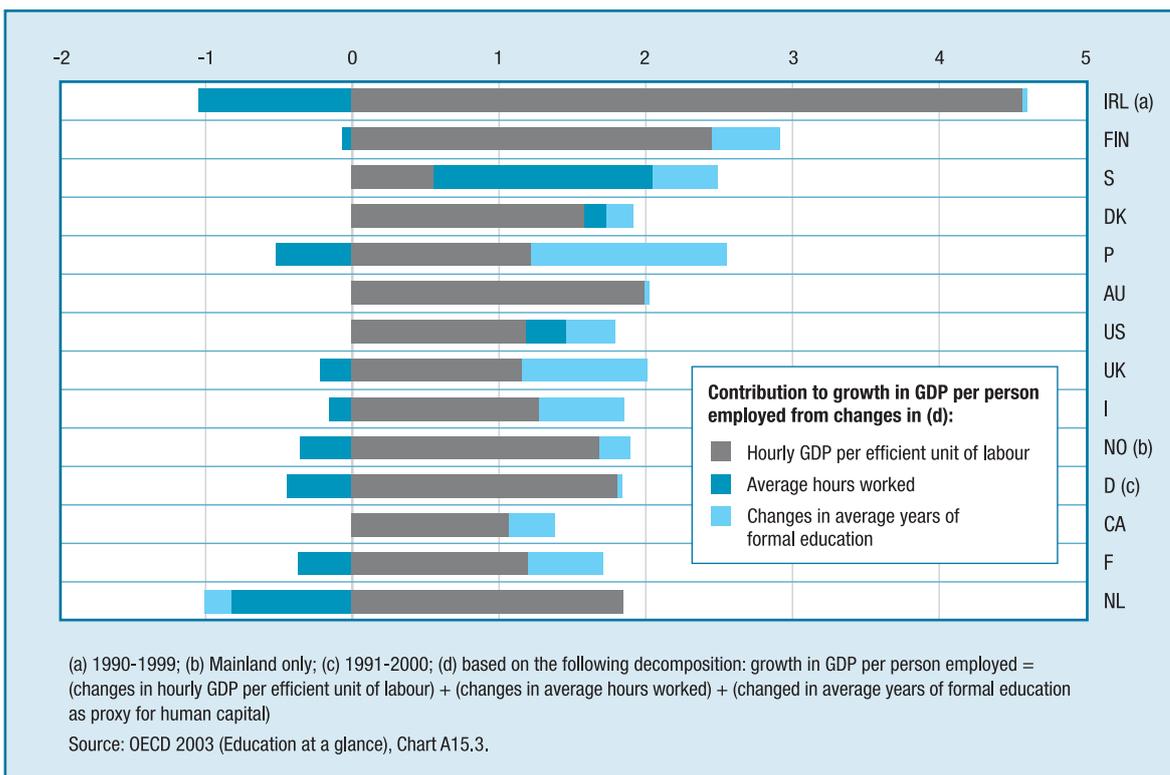


Illustration of impact

A 1 percent increase in the average percentage of working-age population in secondary school is associated with a 0.66 percent increase in GDP per working-age person. A 1 percent increase in human capital stock is associated with a 0.28 percent increase in GDP.

A 1 percent increase in human capital stock is associated with a 0.27 percent increase in GDP per worker. At the sample mean, an increase in average education by one year would raise output per worker by ca. 3 percent.

A 1 percent increase in human capital stock is associated with a 0.57 percent increase in GDP per working-age person. At the sample mean, an increase in average education by one year would raise output per working age person by ca. 6 percent.

An additional year of schooling at secondary and tertiary level is associated with a 0.44 percent increase of GDP per capita. Female schooling at secondary and tertiary levels and male schooling at primary level are insignificant for growth. Female primary schooling promotes growth via lower fertility by 0.39 percent.

considerable uncertainty about the scale of the contribution of human capital to economic growth, the range of estimates being very wide. This also applies to more recent studies using update data sets, which find evidence of positive growth effects.

In Section 1.1, we outlined the two major strands of endogenous growth theory and their implications: those which are based on the stock of human capital (measured mostly by educational attainment or years of schooling) and those which focus on investments in human capital (flows) and mainly use enrolment data in education and training.

For studies focusing on human capital stock, “the empirical literature is however still largely divided on whether the stock of education affects the long-run level or (the) growth rate of the economy. Increasing [the] average [level of] education in the population by one year would raise the level of output per capita by between 3 and 6 percent according to the augmented neo-classical specifications, while it would lead to an over 1 percentage

point faster growth according to the new-growth theories estimates.’ (Sianesi and van Reenen 2002: 41).

The authors conclude: ‘what we consider to be the most plausible estimates in the literature suggest that, *holding other things equal*, an additional year of average school attainment increases the level of aggregate productivity by around 5 % on impact ⁽¹⁸⁾ and by a further 5 % in the long run. This second effect reflects the contribution of human capital to technological progress, i.e. to the development and adoption of new technologies and to the continued improvement of existing production processes.’ (ibid.: 6 f.; italics by authors) The OECD (2003b) has calculated, in a decomposition approach, the contribution to growth (GDP per capita) from changes in:

- (a) average hours worked;
- (b) the average years of formal education (used as a proxy for changes in the quality of labour, i.e. human capital);
- (c) hourly GDP per efficient unit of labour ‘which is equivalent to changes in GDP per worker once changes in working hours and changes in the average quality of labour are accounted for.’ (OECD, 2003b: 173).

The analysis shows ‘that growth in output per employed person is partly attributable to increases in “human capital” of those in employment. [...] OECD countries have invested heavily in education over past decades and this has resulted in a positive contribution of human capital enhancement in growth rates of GDP per person employed and on labour productivity. Over the past decade, skill upgrading amongst workers was particularly marked in Europe, although it was accompanied by sluggish employment growth because productivity gains were achieved partly by dismissals or not employing workers with low skills. [...] education plays an important role in this equation, not only as an input linking aggregate output to the stocks of inputs and technical efficiency, but as a key determinant of the rate of technological progress that affects the output per worker. [...]

⁽¹⁸⁾ I.e. immediately after an increase in years of schooling.

One of the factors behind the good growth record of some countries has been the availability of a large pool of qualified personnel, and skilled labour shortages are rightly considered as a constraint on the growth process.' (ibid.: 173 f.). Figure 4.1 shows the contribution of these effects to GDP per capita growth for some OECD countries.

When 'flows' of human capital are considered, i.e. educational investments measured by school enrolment rates, growth regressions based on endogenous growth theories suggest that increasing overall school enrolment rates by one percentage point leads to an increase in GDP growth per capita of between one and three percentage points (Wilson and Briscoe, 2004).

Sianesi and van Reenen (2002: 41 f.) identify the following additional results:

- (a) in contrast to previous studies '[...] recent contributions based on a new harmonised data set for the OECD, as well as on more sophisticated techniques exploiting the time-series dimension of data have [...] obtained significant and robust estimates in models with high explanatory power;
- (b) the impact of increases in various levels of education appears to vary greatly according to the level of a country's development. In particular, while primary and secondary skills appear to be related to growth in the poorest and in intermediate developing countries respectively, it is tertiary skills that are most important for growth in OECD countries' ⁽¹⁹⁾;
- (c) 'in addition to its direct contribution to growth, human capital has indirect effects as well, by stimulating the accumulation of other productive inputs (e.g. physical capital, technology or health) which in turn foster growth, while discouraging factors (like population growth or infant mortality) which hamper growth.'

1.2.2.1. *Levels of education and training: effects of marginal investment*

What is the contribution to growth of primary, secondary and higher education in countries of different stages of development, and how much should be invested? These are crucial questions on which available evidence allow some conclusions ⁽²⁰⁾:

(a) Effect of primary education on economic growth

In general, and for all countries considered, a one percentage point increase in primary school enrolment rates leads to an around two percentage points increase in GDP growth per capita. The same increase in the stock of people with primary educational level would result in a growth rate of less than one percentage point. Both types of impact are larger for the poorest development countries and not observable for the more advanced OECD countries because primary education is compulsory in the latter countries.

(b) Effect of secondary education on economic growth

A one percentage point increase in secondary school enrolment rates (flows) leads to a 2.5 to 3 percentage points increase in growth; however, the effect is smaller (around 1.5 percentage points or even zero) for OECD countries. As to the stock, one additional year of secondary education apparently leads to a 0.5-1.2 percentage points faster growth, again with no impact for OECD countries. Most studies conclude that secondary education fosters growth primarily in the more advanced developing countries, and less in OECD countries because major parts of secondary education are compulsory here or have already reached a high level.

(c) Effect of tertiary education on economic growth

In OECD countries and according to Gemmel (1996) a one percentage point

⁽¹⁹⁾ We should note, however, that research on the contribution of training to growth is practically non-existent.

⁽²⁰⁾ Based on the review of a number of research studies by Sianesi and van Reenen (2002: 27 f.). Note that the results differ if growth regressions are based on flows (such as enrolment rates) or stocks (such as educational attainment of the population). For a discussion of these different models see Part 2, Section 2.1.1.2.

increase in the annual growth of human capital increases economic growth by 5.9 percentage points. De la Fuente and Doménech (2000) find a strong and robust relationship between output growth per capita and one additional year of tertiary schooling of around 3 %. Bassanini and Scarpetta (2001) even estimate that increasing average higher education by one year would raise output per capita by 6 %. However, the contribution of tertiary education to growth is mostly insignificant, or even negative, in developing countries.

An example of time-series modelling is Asteriou and Agiomirgianakis (2001), who explore the long-term relationship between formal education and GDP in the Greek economy. This study finds a significant relationship between primary, secondary and higher education enrolments and GDP per capita. The main direction of causality runs through the education variables to economic growth but, in the case of higher education, there exists an endogeneity bias in terms of reverse causality: enrolment in higher education is a result of higher GDP per capita and, therefore, of higher incomes of people (parents).

Graff (1996) qualifies this observation by distinguishing between countries with higher and those with lower educational imbalance. The indicator used for identifying imbalances is the ratio between enrolment in higher and enrolment in primary education. This distinction takes account of the fact that in many developing countries higher education is mainly accessible to elite groups of society and does not necessarily reflect economic and social needs. The results tend to confirm that in countries with a more balanced educational system, human capital at intermediate and higher level increases economic growth considerably. In countries with an unbalanced system, however, the contribution to growth of human capital at intermediate level decreases significantly, and that of higher level even becomes negative.

Krueger and Lindahl (1998) find evidence in their data that the relationship between human capital and growth is not linear but decreasing

(inverted U-shape relationship): beyond a certain threshold of years of schooling in a country the impact of schooling to growth diminishes. They find the peak at 7.5 years of education which is below the average education level of 8.4 years for OECD countries. This means that average OECD countries are located at the downward-sloping segment of the education-growth curve which implies decreasing growth returns of a further expansion in education in developed countries. A similar finding is presented by Sianesi and van Reenen (2002). To prevent confusion: the contribution of education and training to GDP may still be positive, but decreases beyond a certain level.

In conclusion, available evidence confirms a positive contribution of education and training to economic growth. However, equally important as these quantitative results are more qualitative links between educational investments and growth. Sianesi and van Reenen (2002) summarise the most important research findings:

- (a) the impact of increased educational levels varies with the stage of a country's development. Primary and secondary education appear to impact on growth mostly in the poorest and intermediate developing countries, whereas tertiary skills are more important for growth in OECD countries, also because primary and major parts of secondary education are compulsory in these countries. Although the direction of causality is unclear – individual demand for education, and its provision, could also be influenced by higher (parental) income and more advanced economic development in a country – ‘both the initial level and the subsequent growth of tertiary education were found to be positively related to per capita income growth in OECD countries’ (ibid.: 35);
- (b) human capital investment has, in addition to its direct effect on growth, indirect effects on economic performance. Human capital may yield additional benefits to growth if it stimulates the accumulation of other productive inputs such as physical

capital, technology, health, crime avoidance and social cohesion (see also Chapter 2). These in turn may foster economic growth. Estimations of these indirect effects use regression techniques similar to those modelling growth (however with a different dependent variable). Following Sianesi and van Reenen (2002);

- (i) human capital appears to be associated with larger investments in physical capital;
- (ii) human capital has a positive effect on productivity growth by raising the rate at which leading-edge foreign technologies are adopted;
- (iii) human capital, and in particular female education, seems to be associated with significantly lower fertility and thus lower population growth, which could increase productivity ⁽²¹⁾;
- (iv) educational attainment is associated with higher life expectancy, lower infant mortality and higher levels of primary and secondary school enrolment rates, which would all lead to higher economic growth if the positive link between education/training and growth holds.

Although research gives some guidance for policy there are a number of caveats in terms of methods and data quality. Needed is 'work reconciling micro and macro evidence by combining data at different levels of aggregation (individual, firm, industry and economy-wide). [...] Improving the (time-varying) measurement of the stock of human capital, taking account of quality issues as well as of further investments in the form of training, should also be high on the agenda. Finally, a more policy-oriented approach would attempt to open up the "black box" of education by trying to explore the mechanisms through which human capital affects growth, for instance by looking at more

disaggregated issues in more detail and in a more satisfactory way than done to date.' (Sianesi and van Reenen, 2002: 42). Closely related to the discussion on the contribution to growth of different levels of education/training (see above) is the question of how efficient are different allocations of financial resources in education and training. This aspect is typically ignored in macro regressions on the sources of economic growth but has important policy implications in terms of the allocation of educational resources. Judson (1998) provides evidence that it is the allocation of educational spending that matters for economic growth more than the level of educational investment. The author defines the efficiency of allocation of educational spending between primary, secondary and tertiary education in a micro model ⁽²²⁾ and concludes that, although for many countries there is a gap between actual and optimal enrolment rates, several economies seem to allocate their educational resources in a nearly optimal way. Those countries that allocate their educational resources inefficiently gain little from their investments in human capital in terms of growth.

1.2.2.2. *The contribution of vocational training to growth*

An additional issue to the research on human capital and growth is the contribution of vocational versus general or academic education. Most growth models that investigate the link between human capital and GDP growth define human capital by proxies and focus on data such as enrolment rates in education or average years of schooling at primary, secondary or tertiary levels. These are mostly, for macro-level analyses and comparative studies, the only data available and include vocational training if acquired in a formal way at vocational schools or by apprenticeship training ⁽²³⁾.

⁽²¹⁾ This link is not always clear-cut. It only applies if a certain level of GDP is maintained and the number of workers (nominator in the ratio GDP/capita) is decreasing thus leading to higher productivity growth.

⁽²²⁾ The ratio of the achieved RoR to the maximum possible RoR the country could obtain given its overall education budget and actual relative costs for each educational level.

⁽²³⁾ ISCED for example includes VET at lower and upper secondary level.

The specific role of (formal) vocational training in economic growth has, however, been scarcely addressed by macroeconomic growth studies. Is it, secondary and upper secondary level, general education or vocational training which contributes more to growth and productivity? Does school- or company-based vocational education and training (VET) (e.g. apprenticeship training, CVT) yield higher impact on economic performance? The answer to these questions would inform where funding should go, who should pay and who benefits ⁽²⁴⁾. At macro level, such analyses would need to isolate vocational training within secondary level. However, we did not come across a study of this kind.

Another issue where we find considerable research gaps is the impact of human capital accumulation through on-the-job training, which is mostly informal or non-formal. Human capital theory predicts that individuals with higher levels of education have greater incentive and are offered more opportunities to accumulate further human capital by continuing training, including on the job. This is confirmed by community labour force surveys and continuing vocational training surveys which include information on continuing training in enterprises (see Figure 4.5 and Table 4.6). Also, microeconomic analyses confirm the prediction that the more highly educated also enjoy enhanced work-related training in their working life (e.g. Blundell et al., 1999).

In this context we should note that research increasingly confirms that both workers and companies profit from investment in employee training, either within apprenticeship training programmes or on-the-job training measures. This issue will be addressed in Chapter 3. Vocational training, and in particular on-the-job training, is closely connected to corporate production strategies and so to national output

growth (Broadberry and Wagner, 1996). However, the variability in vocational training activities between countries makes it very difficult to measure and quantify the impact of different forms of training in macroeconomic cross-country growth models. Wilson and Briscoe (2004) conclude that it is unlikely in the near future that these models will be able to incorporate such training variables.

1.2.2.3. *Quality of schooling*

Other important issues in the discussion on the contribution of education and training to growth refer to the quality of schooling, the allocation of educational resources, the hours worked and the significance of the social infrastructure or social capital for growth ⁽²⁵⁾.

On the question of quality, cross-country growth regressions assume that one year of additional schooling, in advanced economies, is equivalent to one additional year of schooling in developing countries. However, Hanushek and Kim (1995) and Hanushek and Kimko (2000) note that this is a rather crude measure, since education systems vary considerably in terms of quality, for example concerning resources, organisation, duration, etc., and therefore are hardly comparable.

Recent research provides strong evidence that the quality of schooling may be just as important for growth as its quantity ⁽²⁶⁾. These studies include national scores in achievement tests ⁽²⁷⁾ as explanatory variables in standard growth equations and find large and significant productivity effects. Hanushek and Kimko (2000) adjust for differences in education systems by using direct measures of cognitive skills of individuals, interpreted as a measure of schooling outcomes. They use international mathematics and science test cores as direct measures of the labour force quality. They find that school quality – reflected in student achievement tests (IEA and IAEP tests) – has a strong and stable relationship with growth.

⁽²⁴⁾ For a discussion on research results on the significance of education and in particular training for enterprise performance see Chapter 3.

⁽²⁵⁾ The latter issue will be discussed more in Section 1.3.

⁽²⁶⁾ See for example: Lee and Lee, 1995; Barro, 2000; Lee and Barro, 2001.

⁽²⁷⁾ Mostly, international comparative tests of academic achievement are used.

Box 4.1. Impact of the quantity and quality of schooling on growth

Barro (2001) examined roughly 100 countries at different levels of development and in the time periods 1965-75, 1975-85, 1985-95.

The results suggest:

- one additional year of schooling subsequently raises the growth rate by 0.44 % p.a. This implies a social rate of return on secondary and higher education of males of around 7 % p.a. Schooling of females at secondary and tertiary levels is statistically insignificant as is primary schooling of males, but education of females at primary level likely promotes growth indirectly by encouraging lower fertility;
- when test scores are used as an indicator of the quality of schooling, and in particular science test scores in male upper-secondary education, the growth rate would rise by 1.0 % p.a. School attainment in this model is lower and only has an impact of 0.2 %.

Barro concludes that growth is positively related to the starting level of average years of school attainment of adult males at secondary and higher levels. As this is complementary with new technologies, the results suggest an important role for the diffusion of technology. However, growth is insignificantly related to both females at secondary and higher levels and to schooling of males at primary level. Education of women at primary level stimulates growth indirectly by inducing a lower fertility rate.

Comparing the quality of schooling by using test scores on internationally comparable examinations in science, mathematics and reading, it appears that scores, particularly on science tests, have a stronger positive relationship to growth than the quantity of schooling. This suggests that both quality and quantity of schooling matter for growth, but that quality is much more important.

Source: Barro (2001).

Furthermore, when comparing East Asian countries with a lower level of economic performance with the US, they find no reverse causality, i.e. it is not economic growth that causes school quality but the other way round. A more recent study combining standard endogenous growth models and quality of schooling is Barro (2001). He concludes that both the quantity and quality of schooling matter for growth, but that quality is much more important (Box 4.1).

Other proxies for the quality of school resources are pupil/teacher ratios and average teacher salaries. However, only some studies found a positive effect of these indicators on economic performance (Wilson and Briscoe, 2004).

De la Fuente and Ciccone (2002) note another important finding. Most countries with high average pupil performance towards the end of compulsory schooling are also very successful in raising the performance of students from the most disadvantaged backgrounds (OECD, 2001c). Hence, there is scope for education policies that both raise the average quality of human capital and improve social cohesion.

The basic conclusion that can be drawn from current research is that accounting for labour force 'quality' significantly improves the explanation of growth rates. Even though some uncertainties remain as to the appropriateness of various proxies for quality and the actual magnitude of this effect, this finding highlights that both the quantity and quality of human capital have considerable impacts on growth. It is also a key finding that, when controlling for labour force quality, the magnitude of the impact of the number of years of schooling on growth decreases in favour of indicators of quality of schooling.

1.2.3. Education, training and economic performance at regional and local level

Macroeconomic growth models, particularly those based on neoclassical growth theories, are often criticised for neglecting spatial aspects and existing disparities in economic development between regions. Furthermore, the assumption of perfect mobility of physical and human capital, and of frictionless diffusion of technological knowledge, does not correspond to reality. It is argued that knowledge and innovations do not diffuse instantaneously or at an even rate over the economy as a whole.

This refers particularly to advanced knowledge – whether codified or tacit – which is embodied in skilled workers and is mainly concentrated in metropolitan areas. These

inter-regional disparities are in opposition to neoclassical growth theories which assume, via perfect markets and unlimited mobility of production factors, short-term convergence of those imbalances.

1.2.3.1. *Inter-regional inequalities*

The unequal development of skills and human capital are likely to be pronounced spatially, among regions and across countries. For example, Dunford (1997) finds persistent disparities between UK regions in productivity and labour-market participation. A recent study by Huggins and Izushi (2002) also confirms significant variation in productivity among advanced regional economies around the globe.

In countries which are characterised by increasing decentralisation and devolution of responsibilities to regional and local bodies, regions compete against one another, within and across nations, by attracting investments and supporting local businesses to increase their competitiveness in the global market. Many regions, however, and particularly those in peripheries, face persistent regional disparities in productivity and corporate functions, despite the equalising effects of national policies in terms of tax system, teacher salary adjustments and special regional policies (including those at EU level, e.g. ESF and Regional Funds).

Izushi and Huggins (2004) raise the question of whether the infrastructure and provision of education and vocational training at regional level provides some territorial units with comparative advantages over others. To answer this question it is necessary to investigate how investments in human capital influence regional and national economies, and what are the links between the two geographic layers.

1.2.3.2. *Regional development and growth: empirical evidence*

A study by the OECD (2001b) investigates the links between skills acquisition and economic growth in 180 regions across 15 EU Member States. The model correlates primary, secondary and tertiary education levels with regional GDP per capita. The results display that – while tertiary education remains important – the strongest significant association with regional growth is secondary education and training. A higher level of education is critical for R&D and related innovations, but it is secondary education and training that provides the intermediate skills that are critical to industrial know-how and learning-by-doing ⁽²⁸⁾.

Izushi and Huggins (2004) have reviewed literature in economics and economic geography and examined the way in which human capital development facilitates economic growth in European regions in the 1990s ⁽²⁹⁾. Their results show distinct patterns of regional investment in human capital formation. In particular they found a close association between supply and demand for ‘high-order human capital’ (e.g. professionals, R&D workers, executives and other highly skilled workers). Those regions with higher level of investment in tertiary education also display larger concentration of information and communication technology (ICT) sectors – including provision of ICT services and manufacture of ICT devices and equipment – and research functions.

However, in contrast to the OECD study (2001b) mentioned above, no significant association was found between such high-order functions and investment in upper secondary education, both general and vocational types. No consistent patterns were found for regional investment in upper secondary or higher education. However,

⁽²⁸⁾ However, these findings contradict some macroeconomic growth models mentioned above which find, for developed countries such as OECD countries, significant links between human capital and growth, mainly for higher education.

⁽²⁹⁾ Izushi and Huggins used Eurostat data that cover regional units (NUTS-1) in EU countries (Belgium, Denmark, Germany, Greece, Spain, France, Ireland, Italy, Luxembourg, the Netherlands, Austria, Portugal, Finland, Sweden, and the UK). Because of the definition, some countries are included as regions (i.e. Denmark, Ireland, Luxembourg). Further, regions in Sweden, as well as regions in some part of Portugal and Finland, are based on NUTS-2, a lower level of units. However, lack of availability meant the study could not gather all data for all regions.

some regions that include large rural areas have invested in tertiary education at a faster pace than metropolitan regions and thus are closing gaps in the supply of high-order human capital.

The results indicate that the existing stock of human and physical capital in those regions with a high level of urban infrastructure does not lead to a high rate of economic growth. No significant effects of scale have been found that would favour those regions with a larger stock of human capital. Instead, the rate of economic growth is determined by education and training investments, i.e. the accumulation of human as well as physical capital.

The authors conclude: 'the primary policy implication of our study is that, in order to facilitate economic growth, education and training need to supply human capital at a faster pace than simply replenishing it as it disappearing from the labour market. Given the significant impact of high-order human capital (such as business R&D staff in our case study) as well as the increasingly fast pace of technological change that makes human capital obsolete, a concerted effort needs to be made to facilitate its continuous development.' (Izushi and Huggins, 2004: Abstract).

1.3. Social capital and economic performance

What makes some countries accumulate more human capital than others or what makes them more efficient than others in the use of such inputs? Some authors, such as Osberg and Sharpe (2000), argue that GDP per capita is an inadequate indicator of the overall economic well-being of a nation. They put forward that the link between social capital and economic well-being is much stronger than often implied when simple GDP measures are used in growth models. Higher levels of education are typically associated with better environment, higher levels of public health and greater social cohesion, all

of which would be expected to foster economic growth measured in the wider sense. This refers to the notion of 'social capital' as a complement, and also ingredient, of human capital. Despite numerous measurement difficulties, recent research increasingly tries to examine the links between economic growth and social capital ⁽³⁰⁾. Furthermore, an increasing body of research on social capital has been produced by organisations such as the OECD. The majority of this work is grounded in social and political science rather than in economics.

Several studies have constructed quantitative indicators of social capital and related them to economic variables, including productivity. They mainly use an index of trust, derived from the World values survey (WVS) (Box 4.9). Discussion of the links between social capital, social cohesion and trust – which can be seen as macrosocial benefits of education and training – is found in Section 2.1.5. The following section focuses on the links between social capital and economic performance.

Research has mostly confirmed the positive effect of social capital on social ties and institutional quality. Therefore, human capital policies that increase social cohesion by reducing inequality and the social distance between individuals are also likely to improve economic performance (de la Fuente and Ciccone, 2002).

Knack and Keefer (1997) examined various empirical proxies for social capital and assess their impact on economic growth at the country level (Box 4.2). The authors found that trust and civic cooperation are associated with stronger economic performance. Furthermore, they found a strong correlation between trust and average years of schooling. Education is argued to strengthen trust and civic norms, so producing an external effect from investment in human capital. However, this link between trust and economic growth is an indirect one, via education.

Woolcock (2000) and Temple (2001) measure social capital by the extent of trust in

⁽³⁰⁾ See for example Temple, 2000; Osberg and Sharpe, 2000; Woolcock, 2000, and the literature reviews in these papers.

Box 4.2. Social capital and economic payoff – results from the World values survey

Knack and Keefer (1997) investigated 29 market economies from the WVS. Following Putnam (1993a, 1993b), the indicator they used for 'social capital' was associational activity, i.e. membership in formal groups; this was related to the indicators 'trust' and 'civic norms'. To capture the strength of norms of civic cooperation, they construct a variable based on answers to various questions of the WVS about how individuals evaluate anti-civic behaviour.

The authors found that membership of groups is not associated with trust and civic norms. The latter are stronger in countries with higher and more equal incomes, with institutions that restrain predatory actions of chief executives, and with better-educated and ethnically more homogeneous populations.

They conclude:

- interpersonal trust and civic cooperation are associated with stronger economic performance;
- associational activity is not correlated with economic performance (in contrast with Putnam's findings 1993a for Italy);
- trust and norms of civic cooperation are stronger in countries with formal institutions that effectively protect property and contract rights, and in countries that are less polarised along lines of class or ethnicity.

Thus, there is an indirect link between education and economic performance: trust and civic norms are stronger with better educated people and, by improving governmental performance and the quality of policies and public services, are associated with higher economic performance.

This result is quite robust in their sample and corresponds to the findings of Helliwell and Putnam (1999) and Green et al. (2004). However, it remains unclear whether it also holds in OECD countries (e.g. Helliwell, 1996a, 1996b; Zak and Knack, 2001).

Source: Knack and Keefer (1997); see also Box 4.9 for a description of the WVS.

a society ⁽³¹⁾. Social capital seems to be positively correlated with a country's economic success, although the results are not very robust and are much weaker when confined to

Box 4.3. Knowledge, intellectual capital and intangible assets

Wilson and Briscoe (2004) conclude from their literature review that the growing interest in knowledge and intellectual capital – as parts of human capital – is another pronounced development in recent research and policies, both at national and international levels. It builds, amongst other things, on discussions on the role of R&D, intellectual property and intangible assets.

There are some important distinctions between information, knowledge, intellectual capital (IC) and the broader range of intangible assets (Nonaka and Takeuchi, 1995). Information and knowledge are seen as inputs into the accumulation of IC which can be defined as all of the knowledge that has been produced or is capable of producing value. 'In other words, intellectual capital is intellectual material that has been formalised, captured and leveraged to produce a higher value asset.' (Wilson and Briscoe, 2004: Section 7.3).

The growing importance of intellectual capital reflects the shift from manufacturing-based production towards knowledge-based production in which intangible assets such as knowledge and services are increasingly seen as the key drivers of economic growth. Much of this knowledge is created through investment in human capital, with high levels of education and training inputs. For example, the price of many goods exceeds its mere material value many times over and the valuation of many companies is significantly higher than their balance sheet valuations.

NB: A comprehensive overview on publications, methods, best practices, training, research, EU initiatives, etc., on IC is given in: www.intellectualcapital.nl/ [cited 10.5.2004].

Source: Wilson and Briscoe (2004).

the OECD alone. There is also some evidence that trust is positively correlated with educational investment and attainment.

La Porta et al. (1997) find that social capital – and in this context also skills and knowledge – improves government performance, including the quality of the bureaucracy and the judicial system. However, they found only a weak direct association between trust and cross-country growth over the period 1970-93 ⁽³²⁾.

⁽³¹⁾ As collected by the World values survey; see also Section 2.1.5.

⁽³²⁾ Besides Putnam's seminal contribution (Putnam, 1993a, cited in de la Fuente and Ciccone, 2002), there are few empirical studies on the role of social capital at regional level. The scarcity of work in this area is due to the fact that there is very little data on institutional quality at the regional level.

Hall and Jones (1999) investigate wider influences on growth using the notion of social infrastructure. This relates to those institutions and policies that shape the economic environment in which private agents – individuals and firms – make their investment decisions. In a good social infrastructure, all agents capture the social returns of their investments, be it productive activities, capital accumulation, skill acquisition, invention or technological transfer and adoption. As a proxy for social infrastructure the authors combine an index of government anti-diversion policies⁽³³⁾ and an index of openness to international trade.

In conclusion, economic performance seems to be strongly associated with social infrastructure. Similarly, a good social infrastructure appears to be positively correlated with sources of economic performance, including human capital. However, studies are rather disparate in empirically proving this link, not at least due to the difficult operationalisation of many of the ingredients of social capital.

1.4. Discussion

The stimulation of economic growth and employment is one of the principal objectives of public investments in education and training; in addition, increases in growth and employment are likely to facilitate achieving social objectives and avoiding social inequalities. Growth in gross domestic product (GDP) comes from the combination of physical capital, labour and human capital and the investments made in them.

However, the contribution of education, training, skills and competences to economic growth, productivity and employment is not easily measurable. For a start, economic benefits are also generated by factors other than human capital. Second, the abilities, skills and competences of people are only partly visible and measurable: many occupational

skills and competences are acquired through social interaction and through formal and non-formal and informal learning taking place in training institutions, at the workplace, at home and in the social environment. In consequence, it is difficult to attribute and measure the economic benefits of education, training and skills, and of human and social capital. This question has nevertheless generated a vast body of research work.

Traditional neoclassical growth theories treated technical progress and its determinants, for example skills, R&D, innovations, as exogenous factors. Research based on the new theories of endogenous growth instead suggests that one of the major sources for growth and competitiveness is human capital, coupled and complemented with research and development, the accumulation and dissemination of knowledge and use of technologies, particularly ICTs.

This implies that policies to impart ICT-related skills to a broad segment of the population and ensure the adequate supply of technical and scientific personnel needed both for development and for adoption of new technologies would be beneficial for the whole economy. Also, it implies that supporting lifelong learning for all could counteract the depreciation of skills in times of rapid technological change.

A number of empirical studies carried out in the past 10 to 15 years identify skills and related investment in education and training as the key determinants of economic prosperity, which is an important precondition for social cohesion and stability. OECD countries have invested heavily in education and training and this has had direct effect on growth rates of GDP per person employed, on labour productivity. Wilson and Briscoe (2004) conclude on this issue: 'Overall, these growth models demonstrate that higher educational investments have had a significant impact on national economic growth. Broadly, the weight of evidence suggests that a 1 % increase in school enrolment rates has led to an increase

⁽³³⁾ These policies are summarised in the ratings given to a country in terms of: law and order, bureaucratic quality, corruption, risk of expropriation and government repudiation of contracts.

in GDP per capita growth of between 1 and 3 %. An additional year of secondary education, which increases the stock of human capital, rather than simply the flow into education, has led to a more than 1 % increase in economic growth each year.’

However, the range of findings concerning the sources of growth is rather large, depending on the country and skill levels considered, the data available, the period observed, the methodology used and the theories on which research is based. Most studies only provide crude answers to the question of which type of skills, knowledge and competences and which level of education and training have the highest predictive capacity for economic growth.

On the relative contribution of various levels of education, it seems that primary and secondary skills contribute to growth in the poorest and in intermediate developing countries respectively while tertiary skills are the most important for growth in OECD countries. Calculating the implications for growth of achieving the EU target for human resources and a knowledge society has, therefore, to take into account the different levels of advancement across European countries. Research on the contribution of vocational training versus general education is practically non-existent and no conclusions can be drawn so far in this respect.

Another issue raised by research is that the quality of human capital is crucial for economic growth, as it is for individual and company performance. It seems to be even more important a factor than the quantity of human capital as measured by years of schooling or level of education. Raising the quality of education should, therefore, be at the centre of human capital policies. Empirical work points towards some concrete steps to accomplish this objective, but considerable uncertainty remains and more research is necessary to identify the determinants of school performance and student achievement. It is clear, however, that the objective of raising the average quality of human capital does (or should) not stand in contradiction with the objective of enhancing social cohesion.

The observation that research findings display considerable cross-country and regional differences in terms of economic growth and human capital of populations gives rise to a basic question: what makes some countries or regions accumulate more human capital than others or what makes them more efficient than others in the use of such inputs? Social capital and infrastructure can be seen as both independent framework conditions for economic success as well as the outcomes of increased education and training investments. Higher levels of human and social capital, and of social infrastructure, are associated with better environment, higher levels of health, trust and greater social cohesion. All these effects have been found to feed back into faster economic growth. However, studies are rather disparate in empirically proving this link, not least due to the difficult operationalisation of many of the ingredients of social capital.

In conclusion, measures aimed at increasing the quantity and quality of human capital should be an important part of any growth-promoting policy package. This is in line with the Lisbon strategy, which takes on many of the recommendations found in research literature. Implementation of adequate human capital policies ‘appears especially important for those regions of the EU that are lagging behind in productivity and income per capita. It is important to recognise, however, that successful action requires a clear picture of the quantity and quality of regional human capital stocks in order to understand local needs and to identify those policies that are likely to be most effective. For example, it would be important to extend to the regional level recent studies that have tried to assess the skill levels of younger cohorts and of the workforce at large, and to support further research into the determinants of the performance of educational systems. These studies can be a useful input for the formulation of a systematic human resources policy that should be an important part of the EU’s ongoing effort to increase regional cohesion.’ (de la Fuente and Ciccone; 2002: 8).

In addition to its direct effect on growth, human capital may yield additional indirect

benefits if it stimulates the accumulation of other productive inputs such as physical capital, technology, health, crime reduction and social cohesion. These in turn may foster economic growth. Human capital is a rather attractive investment alternative – for the State as well as for individuals and companies – if one considers also non-market returns and wider social benefits in terms of health, crime reduction, trust, citizenship and social cohesion. Research findings on the non-material and wider benefits of education are addressed in the next chapter.

2. Non-material benefits and externalities of education and training

Benefits of education and training are not restricted to material benefits but encapsulate also a broad range of non-material benefits which could be termed as social benefits, non-market outcomes or, in the macro perspective and using the notion of Green et al. (2004), ‘wider social benefits’. Figure 2.2 in Part 2 illustrates different kinds of benefit and their relationship to investments in human and physical capital. Non-material benefits accrue at all levels: individual, company and society. In this chapter we will focus on non-material benefits at macrosocial level ⁽³⁴⁾.

Benefits – material and non-material ones – of human capital investment do not only accrue to the investor but may spill over to other agents (individuals, firms) or to society in general and thus may also favour non-investors, for example in terms of quality of life and social cohesion. These externalities to investments in education and training are mostly difficult to measure empirically. Some attempts will be presented in this chapter.

2.1. Features of macrosocial benefits ⁽³⁵⁾

2.1.1. Attribution of social benefits

Macrosocial benefits extend the focus of analysis to wider units such as the family, community and nation. Macrosocial benefits comprise all non-material benefits that accrue to society and cannot only be attributed to particular members or agents within a society. Here the limitations of an individually-based human capital framework become apparent.

Box 4.4. Examples for non-material benefits and human capital externalities

Some examples for non-material benefits and externalities of education and training are:

- formation of social capital and social infrastructure ^(a) linked with social cohesion, citizenship and trust (interpersonal trust, trust in democracy, institutions, etc.);
- income equality and reduction of poverty;
- political, social and cultural participation; openness to new challenges;
- tolerance towards minorities and other races;
- improved health and life expectancy, also of children;
- reduced violent crime and other forms of criminal behaviour;
- increased awareness of environmental protection;
- better parenting and education of children;
- facilitating the diffusion of innovations and new technologies.

Furthermore, both macrosocial benefits and externalities of education and training investments may promote GDP growth indirectly by creating a favourable social infrastructure or avoiding social costs whereby the money saved could be invested in more productive sectors of the economy. In turn, growth in the economy is likely to facilitate policies and programmes directed towards the integration of target groups and social cohesion.

(a) See Section 2.1.2.3 in Part 2 and Section 2.1.5 in this part.

⁽³⁴⁾ Non-material benefits for individuals are discussed in Section 4.2.1.6.

⁽³⁵⁾ This chapter is mainly based on Green et al. (2004).

Green et al. (2004) distinguish between three primary features of macrosocial benefits of education: their non-identifiability, their relational or positional nature and their effects at system/societal level.

2.1.1.1. *Non-identifiability*

Many benefits of education and training, such as improved health or reduced crime, can be attached to an individual, a group of persons or a company. In contrast, macrosocial benefits cannot necessarily be attributed to an individual person or particular community below the national level (although their impact may be felt at this level). For example, social cohesion may be measurable, or at least proxied by some indicator, but it is not possible to quantify the social cohesiveness of an individual or region. Furthermore, macrosocial benefits are distributed unevenly across potential beneficiaries. Education and training can be defined as a semi-public good as they are, in principle, of benefit to all (Box 4.5).

Box 4.5. **Education and training: public or private goods?**

The question whether education and training are public goods or not is since long disputed. The criteria usually applied to public goods are:

- they are of benefit to all economic subjects (non-rivalry);
- no economic subject can be excluded from these benefits (non-exclusion).

Although education and training feature considerable external and spill-over effects, the criteria above do not fully apply. For example, there are considerable individual benefits and private returns which do not accrue to all members of a society; and individuals may be excluded from some forms of training and continuing training. Therefore, education and training can be defined as semi-public goods which should not be provided and financed only by the state but also by the private sector and individuals.

Source: Melck, 1991: 274 f.

2.1.1.2. *Relational or positional nature*

Some benefits of education and training, such as improved literacy, can be ascribed to one individual alone. However, educational equity in terms of the distribution of educational test scores may be expressed only in terms of relations or comparisons between individuals and thus could be considered as macrosocial in nature.

2.1.1.3. *Benefits at individual and at system level*

The delimitation between micro and macrosocial benefits of education is not always clear. Many economists and social scientists view the macrosocial simply as an aggregation of microsocial or individual benefits. This view mainly goes back to the Austrian School economists who believe that macro aggregates are simply compositions of heterogeneous microprocesses. An antithesis is put forward by theorists such as Lockwood (1992) and Mortensen (1999) who do not reduce macrosocial phenomena to the behaviour of individual actors. In the same line, Green et al. (2004: Chapter 1) argue that societal effects are, at least partly, 'irreducible to individual level phenomena since understanding at the macro societal level may require analysis of the effects of some structures and characteristics which are integral to the collectivity or society itself, and which have meaning only at that level.'

In consequence, some macrosocial benefits, although measured through aggregated individual benefits, are often interpreted as 'genuine' societal benefits. For example, a proxy of trust at societal level may be computed on the basis of answers to the question 'Do you trust people in general?' of the WVS. It should be born in mind when using such proxy indicators in comparative studies that a macrosocial benefit such as societal trust in this example is more than the aggregation of expressed individual trust. It includes historical and cultural norms of trust which, although they might be expressed through on individual responses, reflect the overall trustworthiness of a society, its institutions and representatives and these may be different in other countries.

Box 4.6. Health and environmental externalities

Health

A non-material effect of education and training is the improvement of health and life-expectancy. More educated people are more aware of the potential causes of illness and may practise healthier lifestyles and also choose safer occupations. Societies with higher human capital bring in health and safety laws to protect employees. Grossman and Kaestner (1997), after controlling for per capita income, found that people with more education simply live longer. Better educated men face a lower risk of death from heart disease, and children of better educated women have lower mortality, particularly infant mortality rates (e.g. Feldman et al., 1989), women with higher levels of education being more aware of the dangers to their children's health. Wolfe and Zuvekas (2000) demonstrate that better education also prompts lower fertility rates and smaller family sizes give the children better health and survival prospects, particularly in developing countries. Some studies argue that, at individual level, health benefits can add as much as 40 % to the labour-market return on schooling (e.g. Wolfe and Zuvekas, 1997).

Household management and education of children

Higher education and training leads to a more efficient management of the household and the education of children. For example, households headed by more educated persons achieve higher returns on financial assets. Furthermore, children of better educated parents stay longer and perform better in school (e.g. Solomon, 1975; Angrist and Levy, 1996) and are more efficient learners later in life (e.g. Mincer, 1993).

Environment

McMahon (2000) has argued that higher levels of education will lead to the creation of more sustainable environments, as more informed people are more aware of physical environmental damage. Initially, faster economic growth is associated with global warming, deforestation and many forms of pollution, but education may enable people to take steps to rectify some of this damage without sacrificing the growth gains. Most developed societies with higher human capital and lower poverty have environmental protection laws. The spillover of education to the physical environment is, however, an expanding area of specialist study that lies beyond the reaches of the present report.

Source: Wilson and Briscoe (2004); see this study for literature references.

2.1.2. Non-material benefits and externalities

In addition to non-material benefits of education and training, such as improved health and reduced crime, there are other indirect effects of education and human capital which spill over to other agents or society (externalities) and which macroeconomic models may not be able to capture entirely because it is not always easy to distinguish between benefits and externalities.

Examples of externalities at societal level are social cohesion, support of democracy, social integration, citizenship, income equality and reduction of poverty, innovativeness, quality of life and intergenerational equity or transfers.

Also discussed are the societal effects of individual outcomes, for example:

- (a) the reduction of crime at individual level may improve inner security and may increase the attractiveness of a country for foreign investors;
- (b) improved health of individuals may decrease individual and social expenditure for health insurance; the money saved may be invested in more productive areas and thus improve the economic performance of a country;
- (c) reduced fertility due to better education may delimit population growth and have positive effects on growth, particularly in developing countries;
- (d) higher level of parental education, by improving children's education, may positively influence environmental behaviour and tolerance towards other races or minorities.

However, it should be emphasised that the positive effects of education and training can only be realised if curricula and contents of education and training transmit these values and skills and lead to a sustained change in the behaviour of pupils, students and, later on, adults.

Box 4.6 provides some examples of the externalities investigated.

To assess which non-material benefits at individual level have an effect for society is, however, a rather complex task as there are in

most cases positive and negative side-effects. For example, better health may lead to longer life expectancy. This, however, tends to increase the burden on public and private pension insurance and accordingly reduce the possibilities of investing in other, more productive areas, including education and training. Reduced crime may imply decreasing jobs and increased unemployment of policemen, lawyers, etc. To what extent the positive impacts outweigh the negative ones, and over what time, can only be estimated by a societal bill which includes all effects, direct and indirect ones, and all fields of the economy and society and attempts to calculate the social net returns in a long-term perspective ⁽³⁶⁾.

In conclusion, the impact of education, training and human capital is not only restricted to economic growth but encapsulates a wide area of non-material benefits which spill-over to different agents and society as a whole. Non-material benefits may also have an impact on economic growth which, however, is difficult to measure in quantitative terms. Thus, the direct effects of education on growth as discussed in Chapter 1 have to be complemented by indirect effects or externalities. McMahon (2000) estimates that probably about 75 % of these externalities are non-market outcomes which influence economic growth.

It becomes clear 'that human capital externalities are much more than just a spill-over effect from education in the economy. They are a whole series of net outcomes, most of which are only partially realised after initial impact and many take full effect over a very long time span. These externalities, along with investment in human capital, are important in offsetting diminishing returns to physical capital and so determining positive future per capita growth rates. Investment in education and training is important, not only for its direct contribution to growth but also for the indirect

externalities that it creates, as these eventually feed back into the growth process.' (Wilson and Briscoe, 2004: Section 6.8).

Below, we will discuss in more detail some other examples of non-material benefits and their association with education and training.

2.1.3. Education, training and crime

Green et al. (2004) conclude from their research review that, at the individual level, relatively little quantitative evidence of a direct relationship between education and training and crime could be found when other factors, such as social class, are controlled for. However, there are clearly indirect effects: crime is rather influenced by the overall situation of a country, in terms of poverty and income inequality, unemployment and poor social cohesion. All these antecedents are, however, strongly influenced by education and training which therefore play an indirect role in crime ⁽³⁷⁾.

In addition, in a cross-country comparative context, the effects of education on crime are highly mediated by their national context. For example, some forms of crime are perceived differently between nations (Garland, 2000). Furthermore, we can assume that the macrosocial effects of education on crime become more visible in the longer term (Eisner, 2001). However, relatively little research has been carried out covering longer time spans. Green et al. (2004: Section 2.5) propose some tentative generalisations: 'first, the marginalisation of individuals from labour markets and the norms of society is one of the features of criminal subcultures across European societies, although the form of this marginalisation differs between countries. Second, although there are distinct cultural realisations of crime, there are identifiable common structural antecedents. Although the relationship between unemployment, social disorganisation and crime is unclear, there is emerging evidence that income inequality, and

⁽³⁶⁾ One example of such a societal bill for social protection and sustainable environment has been presented in the framework of the IPTS Future Project (Fahrenkrog and Delgado, 1999). This project produced a collection of papers dealing with several aspects of the European societal bill, covering education, pensions, healthcare and demography, among other things, has been published in 2003 (Zappacosta, 2003).

⁽³⁷⁾ Some findings on these antecedents will be presented in Section 2.2.

Box 4.7. The concept of social cohesion

Following Berger-Schmitt (2000), social cohesion is ‘a characteristic of a society dealing with the connections and relations between societal units such as individuals, groups, associations as well as territorial units’. The sociologist Emile Durkheim who was the first to use this concept, defined social cohesion as the interdependence between members of a society, shared loyalties and solidarity. Features of social cohesion are the strength of social relations, shared values and communities of interpretation, feelings of a common identity and trust among societal members as well as the extent of inequality and disparities (Woolley, 1998; Jenson, 1998, cited in Berger-Schmitt, 2000).

Jenson identifies five dimensions of social cohesion:

- belonging, which implies shared values, identity, feelings of commitment;
- inclusion, which relates to equal opportunities of access;
- participation in political, social and cultural processes;
- recognition, which addresses the issue of respecting and tolerating differences in a pluralistic society;
- legitimacy with respect to the institutions acting as a mediator in conflicts.

Dahrendorf et al. (1995: vii) define a social cohesive society as a society preventing social exclusion: ‘social cohesion comes in to describe a society which offers opportunities to all its members within a framework of accepted values and institutions. Such a society is therefore one of inclusion.’ Other scientists (e.g. McCracken,

1998; Maxwell, 1996, cited in Berger-Schmitt, 2000) underlined that social capital of a society is an essential foundation of its social cohesion.

Berger-Schmitt (2000: 4) concludes that ‘social cohesion incorporates mainly two societal goal dimensions [...]:

- the first dimension concerns the reduction of disparities, inequalities and social exclusion;
- the second dimension concerns the strengthening of social relations, interactions and ties. This dimension embraces all aspects which are generally also considered as the social capital of a society.’

Policies of social cohesion are an important aspect of the efforts of the Council of Europe and other national and international organisations (e.g. Unesco, ILO, OECD) to strengthen human dignity and social rights in a spirit of solidarity. They cover a variety of actions to combat inequalities, promote protection of groups at risk and reinforce supporting measures to family policy.

On a practical level, the Council aims to identify effective models developed in certain countries and introducing them systematically in other countries. These activities are intended to assist the reintegration of excluded persons in five main areas: access to social protection, housing, employment, health care and education. (European Council; <http://social.coe.int/en/cohesion.htm>, [cited 10.5.2004]).

See also the Homepage of DG Social Cohesion: www.coe.int/T/E/Social_cohesion/ [cited 10.5.2004].

by implication education inequality, is an antecedent of some types of crime.’

2.1.4. Education, training and social cohesion

Social cohesion can be a macrosocial benefit of education. It is clearly non-attributable to individuals or groups of persons. It has relational or positional properties in that the positioning of various groups within society is instrumental in producing social cohesion (Green et al., 2004). Social cohesion is also synonymous with social integration and system maintenance, at least in functionalist theories of society (Parsons, 1951) ⁽³⁸⁾.

The degree of a country’s social cohesion (Box 4.7) is historically derived and culturally specific. The role of education and training in fostering social cohesion depends on the distribution of skills and opportunities and, above all, on transmitting related values via education. Although generalisations on the role of education and training across the EU have their limits, there are a number of qualitative comparative studies of vocational education and apprenticeships which explore the mechanisms and levers of value formation.

According to Aldrich (1999), apprenticeship is a training system that has historically involved the social and legal integration of

⁽³⁸⁾ Membership in formal groups was Putnam’s main measure of social capital.

Box 4.8. Etgace – Education and training for active citizenship in Europe

Full name of the project: Education and training for governance and active citizenship in Europe. Analysis of adult learning and design of formal, non-formal and informal educational intervention strategies.

The project was funded by the EU Fifth Framework Programme and run from 2000-02. It covered six countries: Belgium, Spain, Finland, the Netherlands, Slovenia and the UK. The final report was published in 2003 (Holford and van der Veen, 2003).

More information:

<http://www.surrey.ac.uk/Education/ETGACE/>.

youth into society. However, this occurs in quite nationally specific ways. For example, in the UK, an outcomes based system of assessment does not involve the same type of social integration into adult life as the institutional approach favoured by Germany where initial vocational training in the dual system is associated with occupational identity as a part of social integration. Concerning citizenship in a wider sense, the dual system also enables a broad acceptance of youth, supporting their transition into adult life (Evans, 1998).

Nevertheless, highly structured vocational training systems with low flexibility may perpetuate labour-market inequalities. For example, girls and immigrant children typically find training places – and hence later jobs – in only lower status occupations (Bynner, 1994; Brown et al., 2001). This role of education (particularly vocational preparation) in reproducing economic and cultural inequalities is rarely referred to in policy discourses on social cohesion, although it is central to current educational and sociological theory (Morrow and Torres, 1995).

Green et al. (2004) cite a number of European research works dealing with the role of education in reproducing social class. Hatcher (1998) shows that the improvement of social class inequalities through education was rather limited for some countries since the

Second World War. This was also the case in Sweden, a country which is often regarded as an example of egalitarian educational and welfare policy. The pattern of educational inequality remained reasonably constant in Sweden from 1970 to 1990. However, as in other countries, inequality of standards of living has decreased (Erikson and Jonsson, 1996a, 1996b, cited in Hatcher, 1998). This illustrates the inertia of change in class inequalities and the difficulties for educational systems in addressing these inequalities – at least in the short and medium term – and improving social cohesion as an issue purely of increasing educational access.

Still, there are a number of opportunities for education and training to foster active citizenship. The Etgace project (Box 4.8) showed the close links between lifelong learning and various interlocking areas of civic life. However, other comparative studies (e.g. Torney-Purta et al, 2001) show that, at national level, there are no necessary relationships between skills, civic knowledge and outcomes such as social cohesion. Countries scoring most highly on civic knowledge (for example the Nordic and transition countries) are not necessarily characterised by high levels of civic engagement. There is also considerable variation in the relationships between civic activity and other values, such as levels of trust. Again, this indicates that there is no simple causality between education and macrosocial outcomes ⁽³⁹⁾.

2.1.5. Social capital and trust

In Box 2.8 of Part 2, various definitions of social capital were presented. In short, social capital is constituted by the social networks and the norms, values and understanding they share (OECD, 2001d) ⁽⁴⁰⁾ that facilitate cooperation within or among groups (see also Section 1.3).

Green et al. (2004) point out that it is not clear whether social capital is a micro or macro benefit of education. Social capital spans micro, meso and macro dimensions (Lin, 2001;

⁽³⁹⁾ These issues will be further discussed in Section 2.2.1.

⁽⁴⁰⁾ See this publication for a comprehensive discussion and research review on the concept of social capital and trust.

Box 4.9. The World values survey

The WVS is a worldwide investigation of sociocultural and political change. It has carried out representative national surveys of the basic public values and beliefs in more than 65 societies on all six inhabited continents, containing almost 80 % of the world's population. It builds on the European values surveys, first carried out in 1981. A second wave of surveys, designed for global use, was completed in 1990-91, a third wave was carried out in 1995-96 and a fourth wave took place in 1999-2001. This survey has produced evidence of gradual but pervasive changes in what people want out of life, and the basic direction of these changes is, to some extent, predictable. The survey has given rise to more than 300 publications, in 16 languages.

This project is being carried out by an international network of social scientists, with local funding for each survey (though in some cases, it has been possible to raise supplementary funds from outside sources). In exchange for providing the data from interviews with a representative national sample of at least 1 000 people in their own society, each participating group gets immediate access to the data from all of the other participating societies. Thus, they are able to compare the basic values and beliefs of the people of their own society with those of more than 80 other societies. In addition, they are invited to international meetings at which they can compare findings and interpretations with other members of the WVS network.

The project is guided by a steering committee representing all regions of the world. Coordination and distribution of data are based at the Institute for Social Research of the University of Michigan, under the direction of R. Inglehart. The WVS is complemented by a European value survey (EVS) which is carried out jointly with the WVS.

NB: Website and additional information (WVS organisation, network, methods, questionnaires, findings, publications, international statistical databases, etc.) can be found in: <http://wvs.isr.umich.edu/> [cited 19.7.2004].

Baron et al., 2000) and it cannot necessarily be attributed to one individual. According to Putnam (1993b, 2000) it is clearly ecological in nature.

Some studies have constructed quantitative indicators of social capital and related them to economic variables, including productivity

(e.g. La Porta et al., 1997; Knack and Keefer, 1997; Green et al., 2004). They mainly use an index of trust, derived from the WVS.

Trust may be seen as an outcome of social capital or as a component of shared values and norms. 'A distinction can be drawn between: whether people trust others; and whether people are trustworthy [whereby] trustworthiness describes behaviour which results from a multitude of factors including networks and shared values and norms. Three types of trust need to be distinguished:

- (a) inter-personal trust among familiars (family, close work colleagues and neighbours);
 - (b) inter-personal trust among "strangers"; and
 - (c) trust in public and private institutions.'
- (Green et al., 2004).

The second and third wave of the WVS (1990-91, 1995-96, 1999-2001) ⁽⁴¹⁾ showed large differences in reported levels of trust across countries (Table 4.3). The Nordic countries, the Netherlands and Canada rank highest (more than 50 % of respondents said that most people can be trusted), whereas, of the OECD countries, Spain, Mexico, Hungary, France and Portugal (in descending order) score lowest with percentages below 30 %. In Turkey, only 6.5 % of people trust other people.

Concerning the determinants of trust, Knack and Keefer (1997) use the WVS for a sample of 29 market economies. They argue that trust and norms of civic cooperation are stronger in countries with formal institutions that effectively protect property and contractual rights, and in countries that are less polarised according to social class or ethnicity. They conclude that social capital – using indicators of trust and civic norms – matters for measurable economic performance. However, membership of formal groups is neither associated with trust nor improved economic performance. Trust and civic norms are stronger in countries with higher and more equally distributed incomes, with institutions that inhibit predatory actions of chief executives and with better-educated and ethnically homogeneous populations.

⁽⁴¹⁾ Results of the fourth wave (1999-2001) were made available in April 2004 and came too late for this report. First results of the fourth wave have been published by Inglehart, et al. (2004).

Glaeser et al. (2000) use survey and experimental data to identify the determinants of trust and of trustworthiness. They find, amongst other things, that a smaller social distance between individuals, for example due to joint group membership or the same 'race' or nationality, increases both trust and trustworthiness. Alesina and La Ferrara (2000b) identify five broad factors influencing how much people trust others (or not):

- (a) individual culture, traditions and religion;
- (b) how long an individual has lived in a community with a stable composition;
- (c) recent personal history of misfortune;
- (d) the perception of being part of a discriminated group;
- (e) several characteristics of the composition of one's community, including its racial and income heterogeneity.

2.2. Evaluating macrosocial benefits: an empirical assessment ⁽⁴²⁾

Research on the macrosocial benefits of education and training is allocated somewhere between modelling and evaluation. Comparative research on macrosocial benefits cannot be done using summative evaluation methods as

these effects can not be modelled in such a way as to remove confounding influences. Green et al. conclude that 'countries are historically produced, open systems, and it is not possible to subject education systems or nation-states to control or experiment (at least social-democratic ones)' (Green et al, 2004: Section 5.1). It may be possible, however, to increase our understanding of developments and to identify similarities and differences between countries. In doing so, a hybrid approach including modelling and formative judgements may be most suitable.

Green et al. (2004: Section 6) underline that when examining the impact of education and training on macrosocial benefits, there is no single technique to capture the range of benefits, their qualitative dimension and the historical and cultural context in which they are embedded. More suitable are mixed methods, involving qualitative and quantitative techniques to explain the trends involved and referring to factors other than variation between individuals. They investigate the macrosocial benefits by formulating three hypotheses:

- (a) 'macrosocial indicators do not form a coherent syndrome at the national level;
- (b) distribution of education is as important as educational levels in influencing macrosocial benefits;

Table 4.3. The level of interpersonal trust in OECD countries, 1990-91/1995-96

Percentage of people saying that most people can be trusted ^(a)

Norway	65.3	Germany	41.8	Czech Republic*	30.3
Sweden	59.7	Switzerland	41.0	Spain	29.8
Denmark*	57.7	Australia	39.9	Mexico	28.1
Netherlands*	55.8	United States	35.6	Hungary*	24.6
Canada*	52.4	Italy*	35.3	France*	22.8
Finland	47.6	Belgium*	33.2	Portugal*	21.4
Ireland*	47.4	Austria*	31.8	Turkey	6.5
Japan	46.0	United Kingdom	31.0		
Iceland*	43.6	Korea	30.3		

^(a) The question in the survey was: 'Generally speaking, would you say that most people can be trusted, or that you cannot be too careful in dealing with people?'
* 1990-91 data

Source: OECD (2001d) based on the WWS and Knack and Keefer (1997).

⁽⁴²⁾ This section is mainly based Green et al. (2004).

Box 4.10. Education and social cohesion: indicators, countries and methods of analysis

The empirical analysis carried out by Green et al. (2004) at aggregate level includes the following indicators for social cohesion, education and income distribution:

- Indicators of social cohesion
 - general trust
 - associational membership
 - trust in democracy
 - attitude to cheating on taxes
 - attitude to cheating on public transport
 - violent crime
 - risk of assault in the local community
 - tolerance
- Educational indicators (prose literacy scores)
 - prose scores for those with less than upper secondary education
 - prose literacy scores for those with upper secondary education
 - prose literacy scores for those with tertiary education
 - educational inequality: test score ratio 'score of those with tertiary education to those with less than upper secondary education'

- Income distribution

The authors use the Gini ^(e) coefficient as indicator of income inequality.

The following countries are included in the analysis: Australia, Canada, Belgium, Denmark, Finland, Germany, Ireland, the Netherlands, Norway, Poland, Portugal, Sweden, the UK and the US.

Data sources used include:

- World values survey (WVS)
- International adult literacy survey (IALS)
- Interpol crime statistics
- International crime victimisation survey (ICVS)
- OECD statistics

^(e) The Gini coefficient is a measure of relative concentration of a characteristic (e.g. earnings) within a population: how many percent of workers are located in the lower/upper segments of the entire earnings volume of a country? It measures the (double) area between a concentration curve (Lorenz curve) and an equal distribution line. The Gini coefficient takes values between 0 (no concentration = equal distribution) and 1 (full concentration).

Source: Green et al. (2004).

(c) education/skill has a differential impact on various macrosocial indicators according to country context.' (ibid.: Section 6).

To explore these hypotheses, the authors first analysed country level aggregates, and then compared countries using microsocial data.

2.2.1. Social cohesion and macrosocial indicators

The first analysis carried out by Green et al. (2004) investigates whether there is a correlation between several macrosocial indicators for social cohesion and then between educational aggregates and income distribution. Indicators, countries selected and methods used are listed in Box 4.10; the results are presented in Table 4.4.

Correlations across countries between educational level ⁽⁴³⁾ and social cohesion indicators show, at the 5 % significance level ⁽⁴⁴⁾, no significant correlations across countries between education levels. Only at the 10 % level could the authors find a significant correlation with some measures for social cohesion such as 'tolerance', 'never cheating on public transport' and 'risk of assault in the local community'. These results are not surprising because national cultural and institutional factors are likely to outweigh gross education effects on social cohesion.

Given these results, the analysis has been refined. Comparative research suggests that social cohesion is highly sensitive to distributional effects. This refers, for example, to the extent of education and income inequality.

When correlating the indicator for education inequality with macrosocial indicators of social cohesion (Table 4.4) the authors find, at 5 % level, a significant negative correlation between education inequality and the general level of trust. Hence, the higher the level of education inequality, the lower the level of general trust. Those countries with low education inequality (such as Denmark,

⁽⁴³⁾ Here: prose literacy scores for those who have attained upper-secondary education.

⁽⁴⁴⁾ In empirical analyses based on sample data, the level of significance denotes to which degree the observed differences in the distribution of characteristics between the sample and the population are due to random sample errors or are significant. In a statistical test, an appropriate level of significance (e.g. 5 % or 10%) will be predetermined. If the probability for a difference is larger than the chosen significance level, the difference is assumed to be random (i.e. not significant).

Norway and Sweden) also have high levels of trust and those countries with high education inequality (such as Portugal, the UK and the US) have low levels of trust.

It is likely that education inequality is associated with income inequality, although the direction of causality is not really known (45). This being the case, education distribution in a country would be indirectly associated with macrosocial benefits via income distribution.

Therefore, the authors first tested the effects of education distributions on income distributions using Gini coefficients which are a general measure of earnings inequality for the whole population. They found a clear and strong association between distribution of literacy skills and income inequality in the 15 countries compared. Countries with a high degree of skill disparity also have high degrees of income inequality and vice-versa.

In the next stage of their analysis they tested whether there is an association between income inequality and indicators for social cohesion (Table 4.4). In order to consider the general level of economic activity, the results are controlled for the ratio 'GNP per capita' (see

Box 2.4 in Part 2 method). The correlations of income inequality with several indicators for social cohesion are significant. Higher income inequality is associated with a lower level of general trust, higher violent crime and a higher perception of risk of crime in a local community. However, there is also a significant (positive) association between income inequality and civic participation. This could signify that in countries with unequal economic distribution, engagement in groups outside the mainstream channels (e.g. trade unions, political parties) is more important than in countries where income is more equal (e.g. Nordic countries), where civic participation is highly institutionalised.

Green et al. (2004: Section 6.5) conclude 'that there are no necessary correlations between macrosocial outcomes at the national level. [...] The mean level of skill in literacy in each country does not correlate with any measure of macrosocial welfare, although the distribution of this skill does correlate with general trust. The distribution of this skill is also correlated with income inequality which in turn seems to be negatively correlated with many macrosocial benefits including crime.'

Table 4.4. **Correlation of education and income distribution with social cohesion at aggregate level (a)**

	Macrosocial indicators for social cohesion							
	General trust	Civic participation	Trust in democracy	Cheating on taxes	Cheating on public transport	Violent crime	Tolerance	Risk of assault in local community
Mean level of upper secondary attainment in prose literacy	0.354	-0.120	0.244	-0.376	-0.487*	-0.55	0.491*	-0.505*
Educational inequality (b)	-0.592**	0.333	-0.283	0.265	0.171	0.398	-0.060	0.404
Income distribution (c) (Gini coefficient)	-0.562**	0.595**	-0.032	0.430	-0.004	0.660**	0.270	0.628**

(a) Pearson correlation coefficient – (b) Ratio of prose literacy at tertiary educational level to prose literacy at lower than upper secondary educational level in the countries compared (high ratio: indicator of educational inequality) – (c) With control for GNP per capita.

* Correlation is significant at the 10 % level (2-tailed) – ** Correlation is significant at the 5 % level (2-tailed).

Source: Green et al. (2004).

(45) Green et al. (2004: Section 6.4) argue in this context that 'it is quite probable that income equality impacts on educational equality through equalising access to education. It is also likely that social cohesion and solidaristic cultures and political ideologies promote both income equality and educational equality through equalising aspirations and supporting certain types of policy interventions.'

2.2.2. Assessment of macrosocial benefits using microsocial data

In a second approach to assessing the effects of education on macrosocial outcomes, Green et al. (2004) use microsocial data within a cross-country comparative context, addressing similar questions to the macrosocial analysis. First, they examine whether the composition of various social indicators differs between countries. Second, they examine the effects of education level on these social indicators. Some features of this analysis are given in Box 4.1.1.

Figure 4.2 shows, for the countries selected, the effect of education on the four microsocial indicators of social cohesion: political action, institutional trust, support for democracy and race tolerance. The effect of education was controlled for socioeconomic status and age. Each arrow shows the effect size (regression coefficients) ⁽⁴⁶⁾. Significance tests were used to show where the effect sizes differed significantly between countries.

The analysis shows that the effects of education for Sweden and Poland are significantly different from those in the other countries. The effect of education on both trust in institutions and support of democracy was stronger in Sweden whereas the effect of education on political action was lower. The effect of education on race tolerance was higher in Poland than in the other countries. Support for democracy in Western Germany was the lowest of all countries compared, whereas Germans have the highest level of political action.

Green et al. (2004) point to two tentative explanations of these effects:

- (a) one reason could be that the effects represent differences in country means. In countries with a low mean score, for example for racial tolerance, education will have a stronger effect as there is more room for potential increases in the tolerance score. The same may be true for institutional trust, as, for example, in West Germany, Poland and Spain.

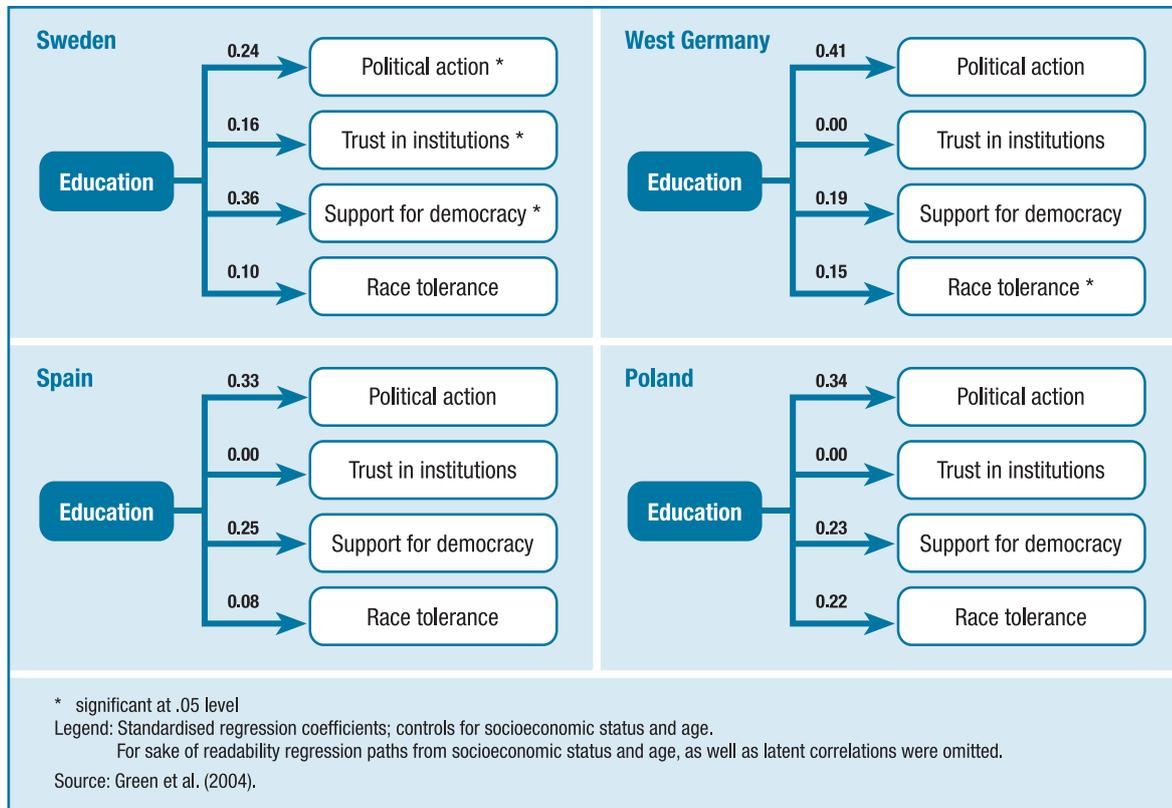
Box 4.11. Data sources, countries and indicators for the analysis of social benefits using microdata

- main data source is the third wave (1995) of the WVS (see Box 4.9);
- the analysis covers four countries: Sweden, Western Germany, Spain and Poland. In total there were 4 390 respondents;
- four 'social outcome' are analysed variables at micro level:
 - 'political action' as a measure of the extent to which people have carried out or would potentially carry out actions directed against 'the establishment' to improve the current state of affairs;
 - 'trust in institutions' was measured by responses to the question 'could you tell me how much confidence you have in {institution}?' This covered three institutions: the legal system, parliament and civil service;
 - 'support for democracy' was measured by the extent to which respondents agreed with the three statements: (1) in democracy, the economic system runs badly; (2) democracies are indecisive and have too much squabbling; (3) democracies aren't good at maintaining order;
 - 'race tolerance' was measured by three items based on the question of which neighbours the respondents would not like to have: (1) other races; (2) immigrants; (3) Muslims.
- the main predictor of these variables is educational level, while the effects of socioeconomic status and age were controlled for. Educational level and status have been standardised to ensure comparability between countries. Initial analyses showed that age, socioeconomic status and education level were interrelated in two ways. First, older participants within each country had a lower level of education and second, educational level and socioeconomic status were positively correlated;
- the effect of education on each of the social outcomes was analysed using the structural equation modelling (SEM) technique. SEM allows analysis of the impact of education on various latent variables measuring the social outcomes simultaneously and comparison of effects to identify systematic differences between countries.

Source: Green et al. (2004).

⁽⁴⁶⁾ For sake of readability regression paths from socioeconomic status and age were omitted.

Figure 4.2. Effects of education on social cohesion in four countries



(b) a second tentative reason put forward by the authors could be ‘that differences in education effects represents differences in the socialisation effects of education [...] in these countries. In Sweden, for example, it could be argued that education has a central role in imparting trust whereas socialisation into forms of political action may occur through involvement in the community or workplace (given Sweden’s high rates on unionisation and the importance of unions in national policy determination). In Poland, the education system may play an important function in terms of increasing racial tolerance, whereas for the other three countries the family or community may have a more important role.’ (Green et al., 2004: Section 7).

The analysis shows that microsocial data may illuminate, although not explain, the

macrosocial effects of education, in particular the role of education in the socialisation process of young people. However, further analysis and more qualitative studies are required.

2.3. Discussion

The effects of education, training and skills on macrosocial outcomes including crime, social cohesion, citizenship, civic and political participation can be analysed through both macrosocial aggregates and microsocial data. Across EU countries, some generalisations can be made concerning the macrosocial benefits of education, training and skills ⁽⁴⁷⁾.

For some macrosocial benefits there are common antecedents, i.e. other influences which themselves are closely associated with education and skills. For example, criminal

⁽⁴⁷⁾ See for the following conclusions Green et al. (2004).

activity and low tolerance could be associated with unemployment, poverty and alienation – which themselves are closely correlated with education, training and skills. This is confirmed, for example, by research on football hooliganism (Dunning, 2000), juvenile delinquency and hate crime (Watts, 2001) for which unemployment and alienation, both related to education, are antecedents. Equally, income and educational inequality are structural antecedents of crime (Kelly, 2000; Lee, 2000).

Using macrosocial aggregates, Green et al. (2004) find clear relationships between education inequality, income inequality and general trust, crime and feelings of community safety. The authors conclude: ‘the implications for education systems are that generalisations concerning the role of education in the rebuilding of civil society [...] or in fostering widespread political involvement through civics education are not applicable across nation-states. While such policies are not necessarily misguided, in that some individuals may benefit, the effect on national levels of social outcomes may be small or non-existent. As our literature review has shown, there is much evidence that increases in the general level of education have not had any effect on national levels of tolerance, crime or social cohesion.’ (Green et al., 2004: Section 8).

The perception of macrosocial benefits such as trust, crime and tolerance highly depends on the societal norms and inequalities in a given country which are hard to change. In the longer term, this change could be fostered by education or training and other institutions involved in state formation (Eisner, 2001). Thus, education and training have an important role to play on the formation of values and the removal of inequalities.

Green et al. (2004) illustrate this point by taking two examples: the Nordic countries and the UK.

Nordic countries are, in general, characterised by high trust, low crime but moderate levels of civic participation. In the

Danish case, this goes along with high levels of lifestyle permissiveness but rather low levels of tolerance towards foreigners. The high levels of trust may be associated with the strong welfare states and historically high levels of ethnic/cultural homogeneity (Knack and Keefer, 1997). However, this may also be a detriment to social cohesion in increasing ethnic heterogeneity. High levels of trust may also relate to relatively high levels of income equality. In this context, education equality may promote trust and lower crime through its impact on income equality.

In Sweden, the strong effect of education on trust in institutions and in democracy may be attributable to social solidarity principles imparted in curricula and in the universal non-selective nature of the primary and secondary school systems. However, evidence that education has a weak effect on civic association in this country may ‘reflect the fact that civic participation in Sweden is highly institutionalised, not least with the prominent role of trades unions within the social partnership system.’ (Green et al., 2004: Section 8).

As a counter-example, the UK has high levels of crime and – compared to the other countries analysed – low levels of both trust and also of tolerance (Halman, 1994). Higher crime and lower trust may reflect the high income inequality (amongst the highest in the EU), and higher levels of intolerance may partly be due to the high immigration over the past 40 years (Halman, 1994) ⁽⁴⁸⁾. Education may play a part in generating lower levels of trust and higher crime through its impact on income inequality. A highly marketised and competitive system as the UK with high levels of inequality in outcomes between schools and regions, and consequently educational inequalities, may generate income inequality and lower trust; this is the opposite of the Nordic emphasis on social solidarity in the school curriculum.

In conclusion, in a cross-country comparative viewpoint, there are specific

⁽⁴⁸⁾ However, this statement should be qualified as other countries, for example Germany, had the same or even higher level of immigration during the past decades.

historical conditions which influence the relationship between education, training, trust, tolerance and social cohesion more generally. Following Green et al. (2004), education and training may have important effects on many of these outcomes, provided that transmitting values such as support for democracy and race tolerance are an important part of the school curricula. However, these effects are mostly indirect and conditional on other – often more powerful – contextual determinants. Much of the work of explaining complex interactions will require more in-depth comparative analysis.

These general conclusions not only suggest that the macrosocial benefits of education and training ‘are rooted both in the distribution of educational outcomes and in the values transmitted through education systems. They are also contingent on the relationship between education and the labour market and other parts of the welfare state. Although there are cultural limits to the extent to which policy borrowing is appropriate with regard to education systems, there are clear lessons for policy-makers. In particular, raising educational, skills and training levels is neither a necessary nor sufficient condition for promoting macrosocial benefits. However, improving the distribution of educational outcomes may be one way in which education and training can make some contribution to more general economic and social redistribution.’ (Green et al., 2004: Section 8).

3. The significance of education and training for company performance ⁽⁴⁹⁾

Research on the impact of education, training and skills/competences on company performance has made some important advances in recent years. There is increasing evidence that employers benefit from human capital and investments in human capital, i.e. mainly training.

The observed wage differentials for workers with different skills paid by firms suggest a clear link between human capital and productivity which is one major element of a firm’s performance. From an economic point of view, wages represent, in theory, the marginal productivity of a worker; a more highly qualified worker can expect higher wages because of (real or assumed) higher productivity compared with lower qualified people.

However, and as discussed in Part 2 Section 2.2, in reality wage scales in companies, collective bargaining, wage rigidities and legal regulations may prevent the pure economic mechanism to come into force.

In general, the association between skills/qualifications and wages as well as positive returns on education and training have been confirmed by numerous studies at the individual level (see Sections 4.1 and 4.2). Thus one issue discussed in this chapter is whether this assumed relationship between skills and performance can also be empirically demonstrated at enterprise level.

3.1. General or specific training: who pays, who benefits?

From a company perspective, investment in human capital (in particular by initial training, continuing training and on-the-job training) bears the risk that trained people may leave the firm and that non-training firms accrue the resultant benefits. According to Becker’s theory (1962) this is particularly the case if firms provide more general training that can also be used by other companies: it is marketable. In this logic, the company has to finance specific training, since this kind of training will only benefit this company, but not general training because it may also benefit other companies. This kind of training should be financed by the worker, paying either directly or indirectly (e.g. by accepting a wage below his or her productivity) for the costs of training.

⁽⁴⁹⁾ This chapter builds mainly on Hansson et al. (2004).

According to Barron et al. (1999), general training represents the majority of company training, about 60-70 % (in the US). In addition, Loewenstein and Spletzer (1999) argue that the generality of training increases with the emergence of more complex jobs. This suggests that most of the company training is also useful to other firms.

It raises the following questions: what the benefits are, who benefits and who should pay for this type of training? Providing answers requires data on a firm's benefit, in particular productivity increase.

The absence of measurement of individual productivity has caused most studies to resort to data on wages as a proxy for productivity. Earlier empirical studies on wage profiles carried out in the 1990s confirmed the prediction of human capital theory on general and specific training: the benefits of a company (productivity) tend to weaken when increasing general training. However, there are other explanations which predict an upward sloping wage curve, such as job-matching models, self-selection models and implicit contracting models⁽⁵⁰⁾.

There is clear evidence that specific on-the-job training increases productivity at the firm level (e.g. Bartel, 1991; Black and Lynch, 1996). Moreover, on-the job-training is a source of innovation and, therefore, long-term competitiveness of firms (e.g. Blundell et al., 1999). However, evidence on the profitability of a firm is mixed, with some studies arguing that profitability increases and others that profitability is unaffected⁽⁵¹⁾. This is not surprising, as the theoretical link between productivity growth at the firm level and profitability is complex.

Still, a number of empirical studies could not support the prediction of Becker's theory that firms benefit mainly from – and therefore only pay for – specific training. Veum (1995) after studying data from the US national longitudinal

survey of youth (NLSY)⁽⁵²⁾ concludes that firms also pay for general training. Equally, more recent studies found that employers do pay for general training because they are able to obtain some of the returns from general training investment⁽⁵³⁾.

The observation that firms (also) pay for general training is important because it suggests that firms benefit from all types of training investments – specific as well as general training. To test whether individual workers contribute to training investment indirectly by receiving wages below their productivity, as suggested by Becker's theory, performance data such as productivity or profitability are necessary. However, data that connect training with productivity or profitability are scarcely available. The absence of company data is striking and few recent studies had access to performance data. These studies will be examined in greater detail below.

Wage compression is one of the theories that try to reconcile empirical findings that firms invest in, and also extract profit from, general training⁽⁵⁴⁾. The basic reasoning is that individuals are not paid their marginal productivity and that firms, therefore, are able to extract higher profits from more skilled workers. Reasons put forward are that internal equity (fairness and moral) considerations in firms' pay structure, collective wage bargaining which may lead to flat pay scales (wage compression) and labour-market regulations, all restrain managers from paying employees the full value of their contribution.

There are numerous other explanations of why firms invest in, and are able to profit from, marketable human capital or training. Autor (2001) applies a model in which firms offer general training to induce self-selection and perform screening of worker ability. This model assumes that more able workers select general training to a greater extent than workers with lower ability.

⁽⁵⁰⁾ For a brief review see Hansson et al. (2004).

⁽⁵¹⁾ For example, Bassi et al. (2001) show that firms investing in training pay above normal returns to shareholders. They also emphasise, however, that while this correlation is consistent with a causal effect, it may also reflect that training is a leading indicator of other factors translating into high profitability.

⁽⁵²⁾ Available from Internet: www.bls.gov/nls/home.htm [cited 29.7.2004].

⁽⁵³⁾ For example, Loewenstein and Spletzer, 1998, 1999; Barron et al., 1999; Autor, 2001.

⁽⁵⁴⁾ See for instance Acemoglu and Pischke (1998, 1999); Bewley (1998).

Another explanation why firms extract rent from general human capital investments is that mobility thresholds reduce the ability of the individual worker to capitalise on these training investments. Factors against turnover (mobility) include using back-loaded compensation schemes (e.g. pay-back of training costs) that induce costs for individuals who change employer. Furthermore, workers may have incomplete information about wages paid elsewhere (Polachek and Robst, 1998; Bewley, 1998).

Despite these theories, the rationale that has attracted most of the attention is the information-based explanation of Katz and Ziderman (1990). Information asymmetry between the training firm and a recruiting (raiding) firm about the training provided reduces the potential benefits that a worker with general training can obtain by moving to another firm. Consequently, information asymmetries render general training specific in the sense that the investment is not observable (verifiable) to other firms. However, other possible explanations have been advanced.

A less explored explanation is connected with findings that both employers and employees benefit from training investments. This implies that individuals employed in firms that provide training expect to receive higher wages in the longer run (possibly also as an effect of training on company performance) compared with employees employed in firms that do not provide any training and therefore are less productive in the longer term. The higher wage growth expected in firms offering training is thus a strong incentive for trained workers to stay with a firm which continuously upgrades human capital instead of the risk of ending up with a new employer whose human capital investment strategy and wage growth is unknown.

Another possibility is, for instance, that higher wages and lower mobility, for example in unionised establishments, typically promote training (Booth and Zoega, 1999). This is confirmed by the Cranet survey (Hansson et al., 2004; see also Section 3.3.3.1): The result that personnel turnover does not determine training might be interpreted as an

indication that turnover does not reduce the incentives for a firm to train employees.

Finally, the employer-employee relationship is complex and it might be myopic to focus only on monetary gains. The provision of general training by the employer might also be explained by the fact that such investments indicate good working conditions and that the employer is committed to his employees. This is likely to reduce the probability (threat) of changing employers.

3.2. Timing of training and impacts

The amount of training investment by firms is also affected by the overall situation of the economy. According to Hansson et al. (2004) the general understanding is that economic expansion, when firms increasingly hire new employees, is associated with an upsurge in firm-sponsored training. However, the results of Dearden et al. (2000) and Bartel (1994) imply the opposite: firms train more when production is low (the pit stop theory). A reason that favourable economic conditions do not induce more company training could be that only a small proportion of company training is geared towards new employees. For instance, only 18 % of all training provided by publicly held companies in Sweden was introductory training for newly hired employees (IPF, 2002). Another reason could be that firms cannot dismiss workers so easily or that firms keep workers if the recession period is expected to last only a short time.

Another finding is that the timing of training apparently does not depend on (previous) investments in physical capital (Barrett and O'Connell, 1999). Similarly, the Cranet survey results (Hansson et al., 2004) indicate that the amount of training provided by firms is not dependent on previous profitability. Both results indicate that the decision of a firm on how much training to provide has little to do with whether the firm has done well in the past and has increased its tangible investments and performance.

A second important issue of timing is when the impacts of the training investment can be observed. There is evidence that such effects emerge after considerable time lags. The results and arguments forwarded in Black and Lynch (1996) and Bassi et al. (2001, 2002) based on US data, d’Arcimoles (1997) based on French data and Hansson (2001) based on Swedish data, indicate that the effects of training materialise 1-2 years after the training period. These results suggest that we should measure the effects of training at least one year after the time of the investment and possibly also over a longer time horizon.

However, the question of when the effects of training investments materialise is not yet answered and it would be beneficial for future research to clarify this issue. Whether the productivity effects of training are lagged or not also has implications for conclusions regarding the wage effects of training for individuals. Does the general view hold that wage increases are (also) caused by an increased productivity from training in previous periods? To which period does this refer? Longitudinal or panel data based on combined data sets for workers and company ⁽⁵⁵⁾ could shed more light on this question.

3.3. Skills and company performance

Economic theory predicts that firms pay higher wages according to the marginal productivity of workers (see Part 2, Section 2.2.1). Higher productivity comes from efficiency at work which is influenced by the commitment, experience and skills/competences of workers. These characteristics are necessary but not sufficient for company performance. They will only translate into higher benefits if the work environment and organisation, human resource management and incentive

structures in a company are suited to exploiting the potential of human capital.

Performance measures include:

- (a) productivity which in its simplest form represents the ratio between output (goods and services) of a company (valued/monetised by the price of the output category) per worker (or per working hour) or per capital unit;
- (b) profitability of the company, i.e. a company’s earnings relative to sales or assets;
- (c) market share, i.e. sales of a brand relative to total sales of that product-type;
- (d) market value with indicators such as stock market value (e.g. price of a company);
- (e) innovations and innovativeness with indicators such as patents, dependent among other things on R&D activities.

3.3.1. Impact of education and training on productivity: review of research findings

According to the literature review by Hansson et al. (2004) few studies have investigated the effect of training on productivity or profitability. One of these is Dearden et al. (2000) who estimate the impact of training on productivity based on different estimation procedures. The impact of a 5 % increase in workers trained would be a 31 % increase in productivity if only the raw correlation between training and productivity is calculated, i.e. if the impact of other variables and their contribution to the total factor productivity is not controlled for. If one controls for other variables, the impact is reduced to 2.6 %. A GMM model ⁽⁵⁶⁾ which deals with the simultaneity problem ⁽⁵⁷⁾, however, increases the impact to 4.1 %. These results suggest that a well-specified regression model with adequate controls works quite well even in the presence of simultaneity problems. Given the restricted data set used in the study, the main concern is not a possible overestimation of the impact of

⁽⁵⁵⁾ For a discussion of merged data sets for employers and employees, see Descy and Tessaring (2001a: 225 f.) and Bellmann (2001: 293 f.).

⁽⁵⁶⁾ General method of moment system estimator. The GMM is a method for parameter estimation that can be viewed as a general case of OLS (ordinary least square) or IV (instrumental variable), two stage least square and other estimations; see for more details Hujer et al. (2004a, Chapters 6-8) and Hansson et al. (2004: Section 3.1.1).

⁽⁵⁷⁾ A simultaneity problem is caused by the effect of previous economic performance on the explanatory variables.

Box 4.12. Surveys carried out by the ASTD

The research done at the American Society of Training and Development (ASTD) is an important source of information. ASTD performs annual benchmarking studies on employee development and training.

A clear advantage of the data collected by ASTD is the quality of the information on training. All companies subscribing to ASTD benchmarking survey gather information on types of training, amount spent on training, etc. All measures of training are clearly defined and so companies provide training information that is collected in a similar way. Studies based on ASTD data thus have much less variance caused by measurement errors in training than most other studies. Committing firms in a survey to a common definition and standard of measuring training, and thus to ensuring comparability of results, is important for research dealing with training and performance. The possibility to connect training data with company outcome measures is another important advantage of the ASTD database. Studies based on this database thus have important benefits compared with most studies of company training.

training, but the extent to which the model underestimates the impact.

Possible drawbacks when analysing productivity are that few studies include the costs of training (e.g. direct and indirect costs); or are able to control for heterogeneity among firms; or address questions like the endogeneity of training ⁽⁵⁸⁾. A further problem is the different or vague definitions of training applied by companies in a particular survey; the ASTD in the US coped successfully with this problem (Box 4.12).

Research in Europe has also made important advances on measuring the impact of training on productivity. Typically, these studies include measures of training and education as part of a firm's innovative capacity, but also variables that are assumed to generate innovations and growth such as investments in IT, R&D, technology, capital intensity, etc. Most studies include a number of control variables.

Selected studies and their main results are

listed below; they are based on the compilation of research work by Hansson et al. (2004) and by Wilson and Briscoe (2004). Table 4.5 provides a summary. Additional references on impact research on training at company level can be found in Barrett et al. (1998) and Barrett (2001).

Dearden et al. (2000) examine the direct impact of training on industrial production of firms in the UK. Their results suggest that company sponsored training generates substantial gains for employers in terms of increased productivity. The study controls for unobserved heterogeneity and potential endogeneity of training using different methods, including a GMM system estimation. The estimates consistently show that the impact of training on productivity is about twice as large as the impact on wages, which means that the company benefit is significant and workers pay for training via reduced wages. The results also suggest that formal training has a greater impact on productivity than informal training. In conclusion, the authors argue that treating training as exogenous leads to an underestimation of the returns on training for employers. This is an important observation, since few studies have eliminated the possibility of two-way relationships between training and company outcome variables such as productivity and profitability (i.e. training has impact on productivity and/or productivity influences training investments).

The results of a study by Barrett and O'Connell (1999) for Ireland suggest that general training has a significant positive impact on productivity, whereas specific company training does not show any positive impact. It is remarkable that these results are found in a model which controls for (i.e. eliminates) changes in corporate innovations and introduction of new personnel policies. These two control variables did not show any significant relationship with changes in productivity. The study confirms that training is a major factor in increasing productivity whereas other plausible and normally uncontrolled factors have little or no influence on productivity.

⁽⁵⁸⁾ See Part 2, Section 2.2.2.

Table 4.5. The impact of education and training on company performance – selected studies

Study, country	Database, Survey	Data	Aim/Subject
Dearden et al. (2000) United Kingdom	British labour force survey (LFS), consensus of production (COP)	Incidence of training, informal and formal training, longitudinal – industry level data	Impact of training on productivity and wages
Barrett (2001) Ireland	EU survey and a follow up survey of Irish business	In-company training (CVT), general and specific training, panel data	Impact of training on productivity, impact of general and specific training
Groot (1999) The Netherlands	Telephone and questionnaire survey	Duration of formal training	Impact of training on wages and productivity growth
Hansson (2001) Sweden	Company database	Type of training, training days, competence, education	Impact of training and competence (skills) on profitability and wages
Gunnarsson et al. (2001) Sweden	LFS, employment register, investment survey, etc.	Proportion with different educational levels	Impact of human capital and IT on productivity growth
Nutek (2000) Sweden	Flex-2 survey (telephone and questionnaire survey)	Learning strategies, competence development activities (measured 0-3 scale), level data	The impact of learning strategies on firm competitiveness
Leiponen (1996) Finland	Survey data, compiled by statistics of Finland (15 manufacturing industries)	Educational level, type of education (technical, natural science)	Impact of education on profitability and innovations
Bellmann and Büchel (2000) Germany	IAB 1997 survey	Amount invested in training	Impact of training on productivity
Bassi et al. (2001) United States	ASTD survey, Compustat	Training investment per employee	Impact of training investments on firm (stock market) performance
Krueger and Lindahl (1998) United States	Company personnel record (manufacturing and service company)	Type of training, basic skills education, occupational courses	Impact of training on wages and employee performance
Blandy et al. (2000) Australia	Questionnaire similar to the UK CEP survey by the LSE	Training quantity in hours	Effects of the on-the-job training on productivity and earnings

Outcome measures	Strength / Weakness	Findings
Productivity measured as change in log real value added per employee	Extensive robustness tests and econometric modelling, data on incidence of training (not days in training)	Training has a positive impact on productivity and wages, with a twice as large effect on productivity. Formal training has larger impact on productivity than informal training
Log sales growth	Days in training, panel data, important control variables	General training has a positive impact on productivity, specific training no impact
Estimates (100 % scale) of productivity growth before and after, trained and not-trained	Direct estimates of differences in productivity / estimates of productivity	Average productivity growth about 4-5 times larger than wage growth. Weak connection between who contributes to training investment and who benefits from the training
Profitability, (revenues net of wage and overhead costs)	Direct measure and human capital stock / level data, single employer	The concurrent impact of training on profit is negative and impact on wages positive. The skills/competence of the individual is significantly related to profitability
Total factor productivity	Lagged IT and human capital effects	Interaction term IT and educational level is highly significant indicating that the increase in productivity growth is largely tied to increase in higher educational level
Productivity (value added), profitability (revenues to cost ratio)	Includes a large number of control variables and competing learning variables / level data, weak training data	Competence development activities have a significant effect on productivity and profitability. Larger effects for larger firms. Education is associated with profitability
Net profit margin, innovations such as patents, improvements, etc.	Use first difference, control variables, 2SWLS to handle simultaneity problems and GMM	Educational competence is significantly associated with profitability. Complementarities exist between different general skills acquired in higher education. Innovative firms are more dependent on educational level in generating profitability
Productivity measured by log annual sales	Large number of observations / level data	Initial results indicate a strong association between training and productivity. However, the results are likely to be driven by 'ability'. Authors stress HRM practices as main factor
Stock market returns, income-, Tobins Q-, and sales per employee	Good quality training data, firm performance measures, and control variables	Training investments are associated with next year's stock market performance. Same result for level and change data. Changes in training investments can predict future stock returns
Performance awards, absenteeism, self-reported performance	Homogeneous workers, compare results at two companies/ weak performance data	The work place education programme had a generally weak effect on employee performance measurements, except for perceived performance. The effect of the training was positive but not significant in many cases
Productivity and profitability	The study uses quantitative data on training and performance / small sample size	Profitability is directly related to: a) quantity and quality of training; b) firms paying above market wage rates; c) firms' difficulties in finding suitable employees. No clear picture regarding the impact of training on productivity

Table 4.5. (cont'd)

Study, country	Database, Survey	Data	Aim/Subject
Maglen and Hopkins (1999) Australia	Case study based on interviews with managers and employees	Training expenditure	Evaluation of the returns on training depending on other factors
Doucouliaagos et al. (2000) Australia	Case study	Cost and benefits of training programmes	Application of a four-step evaluation process of training investments

Source: Hansson et al (2004), to be consulted for more details on the above studies

The study by Groot (1999) for Dutch firms showed that although a considerable number of workers contributed to training investment with their leisure time – for average training duration of close to six months – their benefits in terms of higher wages were relatively small: while the average wage growth induced by training was 3.3 %, the average productivity growth as a result of training was found to be 16 %.

A study by Hansson (2001) for programming consultants in Sweden had access to measures such as profitability, amount of training, wages, and each employee's acquired human capital stock (approximated by the individual's competence profile). Whereas employers not only paid for all direct training costs and lost a considerable amount of profit, no evidence was found that workers contributed to training investment by receiving wages below their productivity. However, the findings suggest that firms recovered their investment in programming training in the long run because the programming skills and competences acquired were significantly associated with profitability. Hansson concludes that the investment in 'marketable' human capital is similar to any other investment scheme that firms normally undertake in their business operations, with an initial investment and a payoff in the future.

Gunnarsson et al. (2001) examined education level and productivity growth over a 10-year period (1986-95) for 14 industries in Sweden. Their findings suggest that the increase in the education level explains a large part of IT related productivity growth. The authors conclude that 'measures to promote

increased use of IT should be followed up by measures promoting skill upgrading. Our results actually show that, in general, upgrading skills at a given level of IT (i.e. share of computers in total capital) has a much stronger growth-enhancing effect than increasing IT investments at a given human capital structure.' (ibid.: 44).

A study by Nutek (2000) on different learning strategies in Swedish companies shows that competence development activities are associated with both productivity and profitability of firms. The effect is more pronounced for larger firms. Training is measured in a broad sense by three company activities: planning, learning on the job and proportion trained. This strengthens the interpretation that employers are able to capture some of the returns generated by these productivity effects. The study also shows that the proportion of more highly trained employees is significantly associated with both productivity (value added) and profitability (revenues to cost ratio). This provides a more solid base for inferring that more training generates higher productivity and that firms are able to capture some of these returns.

In a panel study of 209 Finnish companies, Leiponen (1996a) finds a significant association between educational level and profitability. The results also suggest that innovative firms are more dependent on educational competence in generating profit. The author overcomes the endogeneity problem caused by the impact of previous economic performance on the explanatory variables such as training. Without addressing the problem of endogeneity, the

Outcome measures	Strength / Weakness	Findings
Labour productivity	Control variables are only used for qualitative interpretation of the differences	In the most cases training investments had led to a positive returns, which is dependent on HR practices; the bundling of HR policies is crucial; better performers also planned strategically
Calculations of return on instrument (ROI)	Quantitative estimation of returns on training investments	ROI (%) calculated is between 70 and 7 000

results are largely insignificant. This finding indicates again that the problem of endogeneity between profitability and other human capital variables renders the impact less significant. Leiponen concludes that general competences acquired in education, notably higher and post-graduate education, are beneficial for the profitability of the firm.

A study by Ballot et al. (2001; not cited in Table 4.5) examined the impact of human and technological capital on productivity in a panel sample of large French and Swedish firms. Specifically, measures of a firm’s human capital stock were constructed on the basis of past and present training expenditure. The results confirm that such firm-sponsored training, together with R&D, are key inputs to the market performance of firms in both countries.

A study by Papalexandris and associates (2000; not cited in Table 4.5) into Greek firms found that, where training was treated as a continuous, life-long learning process, it had considerable impact on the growth of firms.

Blechinger and Pfeiffer (1998; not cited in Table 4.5) used survey data for the German manufacturing sector to explore the links between employment growth, technological change and labour force skill structures. They found that innovative firms experienced the highest growth rates and such firms tended to employ more highly skilled workers.

The German Institute for Employment Research (IAB) conducts one of the most comprehensive establishment panels in

Europe with 9 200 participating establishments (see Bellmann, 2001). However, IAB analyses so far focus to a lesser extent on the effects of training investments for firms and more on the effects for individuals. An exception is the study by Bellmann and Büchel (2000) on the effects of continuous vocational training on productivity. Their result is based on 3 400 cross-sectional observations from the 1997 IAB survey. The authors apply different models in estimating the impact of training and control for industry, size and employee characteristics. The initial regression results for both parts of reunited Germany indicate a significant association between training investment and productivity. However, the authors argue that this finding is largely a consequence of a selection problem, with individuals receiving training having more ability and with certain firms being more capable of providing adequate training to their employees. The authors stress the importance of strategic HRM practices in generating productivity effects from training investment.

Wilson and Briscoe (2004) draw attention to a different approach to exploring the links between investment in human resources and productivity in organisations. It was developed over the last twenty years by researchers at the National Institute of Economic and Social Research (NIESR) in the UK ⁽⁵⁹⁾. Much of this work is summarised in Prais (1995) and all the relevant papers have been published in full in various editions of the *National Institute Economic Review*. The research approach

⁽⁵⁹⁾ Available from Internet: www.niesr.ac.uk [cited 29.7.2004].

consists of carrying out empirical international comparisons of productivity and associated schooling and vocational training investments between France, Germany, the Netherlands, the UK, and other selected countries. The concern is with the results of different kinds and qualities of human resource investment. Often matched samples of plants and firms from specific industrial sectors have been used to explore the relationship of productivity with human capital.

Various studies carried out by researchers at the NIESR have sought to demonstrate the links between a better educated and better trained workforce and achieved higher output per worker. Investigations were made into manufacturing plants in the UK and their counterparts in other EU countries. The studies looked into the metalworking, woodworking, clothing and food manufacturing industries, as well as selected service sector trades. It was found that across all these industries the acquisition of skills among the workforce was a critical factor in raising productivity, as measured by output per worker. Recent studies of this kind have been carried out by Mason et al. (1999) into banking services and by Jarvis et al. (2002) into the ceramic tableware industries. Commonly, plants in France, Germany and the Netherlands, where the percentage of vocational qualifications was much higher, return significantly higher output per worker than their UK equivalents.

Based on ASTD data (see Box 4.12), Bassi et al. (2001) found that training investments are associated with stock-market performance in the following year. They conclude that changes in training investment can predict future stock returns.

For Australia, Blandy et al. (2000) found a positive relationship between a firm's profitability and the quantity and quality of training offered by the firm, and that the latter is also correlated with other forms of investments. They conclude that the principal reason why training is profitable for firms is

that it increases the productivity of their employees more than it raises their employees' wages.

Another study for Australia is Maglen and Hopkins (2001; not cited in Table 4.5) who updated their own previous study (1999). They used integrated variables such as work organisation, job design, employment practices and other firm-specific variables to explain the return to training. The authors compared enterprises that produced similar goods or services and were similar in size. Their findings suggest that there is no winning set of relationships that could be translated into a series of best practice procedures, but rather that the key factors in optimising business performance are the links between strategic objectives and practice within the enterprise itself. In a comparison between seven firms in four sectors they found that better performers, measured by labour productivity, planned at strategic level, with training an integral part of strategy. In contrast, the poorest performances lacked these characteristics.

Doucoulagos and Sgro (2000) developed a training evaluation model, which they tested on seven Australian firms with longitudinal firm level data, aiming to calculate the return on investment of training, both financially and non-financially. On the basis of the collected data they identified that the return on investment in training programmes varied between 30 and 7 % for seven different training programmes. They used several indicators to estimate performance and benefits, such as energy saving after training drivers or sales increase after the training of store managers.

3.3.2. Small and medium enterprises

The importance of small and medium sized enterprises (SMEs), often defined as firms with less than 250 employees, for the economy and employment is well known: around 99 % of all firms in the EU are SMEs and they employ around two third of all employees ⁽⁶⁰⁾.

⁽⁶⁰⁾ More details on employment and training issues of SMEs can be found in Cedefop's second research report (Trouvé, 2001; Descy and Tessaring, 2001a).

Table 4.6. Provision and costs of training by size of enterprise, 1999

Enterprise size (number of employees)	Provision of any type of training (total)	Provision of CVT courses		Provision of other forms of training	No provision of training
		% of all enterprises	% of total labour costs		
10-19	49	41	1.5	40	51
20-49	67	58	1.4	57	33
50-249	81	75	2.4	71	19
250-499	94	91	2.3	85	6
500-999	96	91	2.7	90	4
1 000 or more	99	98	2.5	96	1
Total	62	54	2.3	53	38

Source: Continuing vocational training survey (CVTS2), 1999 (NewCronos data base).

Although SMEs are characterised by a great heterogeneity across European countries and industries, most surveys in different European countries show that SMEs, compared to larger firms, invest less in vocational training and use less formalised forms of training. The exception is apprenticeship training in SMEs in a number of European countries. According to the Continuing vocational training survey 2 (CVTS2) in the EU Member States (conducted in 2000, reference year 1999), SMEs, and in particular small firms with less than 50 employees, train fewer of their employees and also invest less in CVT (Table 4.6). SMEs spend around 1.5 % of their total labour costs for CVT, larger firms 2.3 % and more. However, there are clear variations across countries and sectors.

Furthermore, SMEs are less likely to have a clear human resource strategy, training plan, or advanced HRM practices. Other reasons are: lack of resources to plan training or find the right training options; lack of internal or external funds; and unfamiliarity with, or resistance to, training (Pukkinen et al., 2001).

In this context it is also of interest to know how important human capital investments are in relation to other forms of intangible and tangible investments to explain their competitiveness and performance. Flexibility,

entrepreneurship, close relations to partners and customers, motivation of the workforce, and the realisation of niche strategies are important strengths of SMEs in comparison to larger firms ⁽⁶¹⁾. Training appears not to be a necessary condition for the performance of many SMEs. Furthermore, they have difficulties in engaging trainers or providing free time for training, given the small number of employees in the firm.

In contrast to their economic importance, few studies exist which deal explicitly with the impact of training investment on firm performance for SMEs. Hansson et al. (2004) conclude from their literature survey that there is no reason to assume that skills or training in SMEs have less impact on company performance.

A similar conclusion is drawn by Pukkinen et al. (2001: 13), based on SME-expert consultations throughout the EU and a literature review: '[...] there is evidence that investment in training impacts positively on, among other things, company productivity, profitability and product/service quality. Training which improves enterprise and individual competitiveness and competence creates better chances for long-term survival. In the long term, support to encourage CVT in SMEs may lead to economic growth and

⁽⁶¹⁾ See for a more detailed discussion on SMEs and entrepreneurship Descy and Tessaring (2001a, Chapters 3 and 4 of Part 3).

Table 4.7. Selected studies on training in SMEs

Study, country	Database/Survey	Data	Sample and size	Aim/subject
Leitner (2001) Austria	Empirical study based on a questionnaire	Importance of training	100 SMEs	Intangibles, strategy and firm performance
Romijn and Albaladejo (2000) United Kingdom	Survey on 50 SMEs (interviews)	Skills for workforce	50 ICT and electronic firms	Internal and external sources of innovation capability in SME
Bosma et al. (2002) Netherlands	Panel survey among 1 100 new business founders between 94-97	General, industry-specific and entrepreneurship-specific investment	1 100 firms	Does human capital investment enhance entrepreneurial performance?

Source: Hansson et al (2004).

further employment. Therefore, the reasons [why] SMEs train less must be explored.'

The selected studies compiled in Table 4.7 suggest the following results:

- (a) Leitner (2001), when investigating the impact of different strategic investments on firm performance for Austrian SMEs, found that the extent of training was one of the few internal factors that had a direct impact on firm revenues. Considering the size of the firm – he studied firms with 20 to 500 employees – he found that this correlation was highly relevant for smaller firms with less than 50 employees. In general, however, smaller firms invested less in training than larger firms.

Furthermore, Leitner discovered a positive relationship between training: the corporate culture and communication within the company. Given the impact of external factors he found that training was especially important for firms that are situated to very dynamic environments. He concluded that, in general, training investments allow firms to perform better than other firms in different competitive and hostile environments in terms of, for example, mature product life cycles, business cycle-dependent life cycles and dynamic business environment.

- (b) Romijn and Albaladejo (2000) investigated

the role of various internal and external sources of innovation capability in small UK companies. Besides other factors such as R&D investments and interactions with research and training institutions, they found a range of internal factors, such as the owners' technical education and their prior working experiences, technical skills of the workforce and training investments to have a significant effect on innovation capability.

- (c) A large body of research exists on the importance of human capital for the success of business start-ups (see also Descy and Tessaring, 2001a, Chapter 4 of Part 3). In these studies formal education, skills and experiences and the talent of the founder are used to explain business success. However, there is little investigation of the role that training of the workforce or teams plays in a company's success. Bosma et al. (2002) studied the value of human capital for start-up companies in the Netherlands. They separated general, industry-specific and entrepreneurship-specific human capital investments of the founder and measured performance according to survival, profit and generated employment. Their main findings are that investment in industry-specific human capital aspects, such as

Method	Control variables	Outcome measures	Strength/weakness	Findings
Correlation and Anova-Analysis	Industry, size, strategy	Earnings, sales and employment growth	No quantitative measurement of training investments	Firms with regular training have higher returns
Correlation analysis	None with respect to training and performance	Innovation capability (index)	Captures various indicators to explain innovation performance, not possible to relate training directly to performance	Skills of workforce (proportion of university trained) have a positive impact on innovation performance
Regressions	Various controls such as industry, gender, labour-market histories	Survival, profit, employment	Human capital is measured only through experiences	Human capital (experiences of the founder) influences the entire set of performance measures

former experience in industry, and entrepreneurship human capital specifics, such as experience in business ownership, contribute significantly to explaining the performance of small firm founders, whereas general investments, such as level of high education, play a minor role.

3.3.3. Human resource management and high performance work systems

Research on managing human resources at both the firm and the industry levels suggest that workers are a key source of an organisation’s competitive advantage. It is the quality of these human resources that determines organisational performance. Research highlights a number of different issues in this context, such as the role played by human resource management in promoting such resources, the competences of management itself, particularly top managers and leadership. This research work emphasises the importance of education and training in often very complex industrial processes (Wilson and Briscoe, 2004).

It is difficult to answer the question of whether it is the combined effect of human resource management (HRM) practices or

whether it is certain practices, such as employee development, that generate good company performance. Outside mainstream economic research, HRM literature reports a range of results concerning the impact of training and human capital on firm and industry performance. Such research bridges the gap between analyses of individual rates of return (RoR) and macro-level studies. It suggests that companies that employ managers, professional and other staff with higher qualifications can expect to achieve better commercial performance. This reflects the belief that people are a key source of an organisation’s competitive advantage and that it is the quality of the management of human resources that determines organisational performance.

3.3.3.1. HRM and in-company training

In general, there is evidence that training has a greater impact when it is undertaken in connection with new HRM practices. Supportive HRM practices such as the existence of a formal training strategy, written commitments to training, methods for analysing training needs and links between training and strategic objectives are important factors in explaining training outcomes ⁽⁶²⁾.

⁽⁶²⁾ See for instance Maglen and Hopkins (2001); Blandy et al. (2000); Baldwin and Johnson (1996); see also the results of the Cranet survey presented below.

Box 4.13. High performance work systems (HPWS)

High performance work systems are organisations that achieve extraordinary results through their effective utilisation of several core characteristics:

- leadership that empowers and involves others;
- a relentless focus on strategy and results;
- open sharing of relevant information;
- borderless sharing of power and responsibilities;
- team-based working design;
- teamwork reinforced through rewards;
- creativity;
- rapid results orientation.

In HPWS employees participate more in company and job-related decisions, concerning, for example, day-to-day issues of reorganisation in the work process. Employees carry out their work in a more autonomous way, communicate during work with colleagues, superiors, experts and, if necessary, with clients. HPWS is realised in numerous variants, concerning self-regulating teams, group work, specific problem solving groups, quality improvement teams or statistical process control. However, without human resource management, HPWS will not function properly. Continuing training, fostering motivation and employment security are some of the ingredients of such HRM.

Successful implementation of HPWS requires an effective combination of new work organisation, participation by employees, motivational incentives and mechanisms to ensure employment security. In this case, both employees and employers will profit from organisational innovations and new communication and decision structures.

HPWS is also known by such names as: socio-technical systems, participative management, self-management, self-directed work teams, empowered work teams.

Sources: Rayner Ass., available from Internet: www.raynerassoc.com/HomeLinks/HPWS.html [cited 10.5.2004]; Friedrich Ebert-Stiftung, available from Internet: www.fes.de/fulltext/fo-wirtschaft/00715002a.htm [cited 10.5.2004].

The impact of HRM practices on firm performance has received considerable attention. Many special issues of management journals are devoted to human resource management practice and firm performance,

for instance *Academy of Management Journal* (39: 4, 1996), *International Journal of Human Resources* (8: 3, 1997), *Human Resource Management* (Fall, 1997), *The Human Resource Management Journal* (Fall, 1999). The argument put forward in this line of research is that advanced HRM practices produce a higher level of productivity. The findings suggest that there is a connection between HRM practices or what is often called high performance work systems (HPWS; see Box 4.13) ⁽⁶³⁾ and firm performance such as sales, market values, market-to-book values, profitability, productivity, etc.

Boselie et al. (2001) ⁽⁶⁴⁾ provide an overview of the findings of HRM research over the last decade. The results with regard to the effects of training on a company suggest:

- (a) training has a positive impact on the different dimensions of the performance of the firm: product quality, product development, market share and growth in sales;
- (b) higher investment in training results in higher profits, market share and investment in the near future;
- (c) higher investment in training results in a lower degree of staff turnover;
- (d) training has a positive impact on the relationship between management and the other employees;
- (e) training has a positive impact on perceived organisational performance.

An important database in this context is the Cranet survey. It is the largest and most representative independent survey of HRM policies and practices in the world and carries out regular international comparative surveys of organisational policies and practices in HRM across Europe (Box 4.14).

The analysis of the Cranet data carried out by Hansson et al. (2004) should be seen as a first attempt to explore this database. Some interesting findings are summarised below (for more details see Hansson et al., 2004):

- (a) in general, the amount of wage bills spent on training does not correspond to the

⁽⁶³⁾ These human resource management practices are sometimes referred to as human capital enhancing systems, or high commitment policies (systems).

⁽⁶⁴⁾ Cited by Hansson et al. (2004). See this paper for references.

- number or proportion of employees trained;
- (b) the proportion of employees trained in a firm appears to depend on whether the firm can afford to train the employees or not. This conclusion is derived from findings that the number of employees trained in a given year is correlated with prior profitability (reverse causation);
 - (c) the amount of wage bills spent on training is not correlated with prior profitability; this indicates that firms can afford training not because of their previous profitability but that it is rather training that generates profit;
 - (d) the existence of training needs analyses for employees and a written training policy are two important factors associated with the number of persons trained at the firm; the existence of training needs analyses is also associated with the amount of wage bills spent on training in the firm;
 - (e) training measures are, in most cases, significantly higher or more frequently used in high performing firms compared with firms that perform below average in a given sector.

An interesting aspect of these results is that the decision to train is largely determined by firm-specific factors. A general picture seems to emerge where factors that are considered as approximations for good working conditions or the 'good employer' are also largely connected with performance measures such as productivity and profitability. In regard to company training it is important to note that firms with better profitability and stock-market performance also spend more on training and train more of their employees compared with firms performing below average in a given sector. Another observation is that firms with better profitability and stock market performance also have more highly skilled employees.

Whether the effect of training in particular can be separated from other HRM or HPWS practices is difficult to answer. Training is mostly incorporated as one factor among others in the larger construct of HPWS. Few studies measure the additional effect of a

Box 4.14. The Cranet network

The Cranet network was established in 1989 with five founder countries (Germany, Spain, France, Sweden and the UK). It is coordinated by the Centre for European human resource management at Cranfield School of Management. The Cranet survey is now the largest and most representative independent survey of HRM policies and practices in the world. The network itself is a collaboration between 34 universities and business schools, which carries out a regular international comparative survey of organisational policies and practices in human resource management (HRM) across Europe.

Cranet has been running the survey since 1990 using standardised questionnaires sent to private and public organisations in different countries. The standardised questionnaire is translated into each member country's language and adapted to the different national contexts (taking into consideration such factors as legislation, labour markets and culture). During each round of the survey, amendments are made to capture new developments. But on the whole, the questionnaire stays unchanged in order to be able to observe developments over time.

The data is collected through a standardised postal questionnaire (with the exception of Greece where interviews are used to gather the information). The questionnaire is distributed to organisations with 200 or more employees and it is addressed to the most senior human resource/personnel specialist in the organisation. The 1999 survey was distributed to over 50 000 organisations and 8 050 responses were received giving a total response rate of 15 %. The willingness of companies to respond was, for example, higher in Scandinavian countries than in Southern European countries. The 1999 survey database includes 8 487 observations from 27 countries (17 European countries). The number of observations varies because not all companies have answered all questions.

It is important to note that the provider of the information in this survey is the firm. As noted by Barron et al. (1997) employer based surveys typically report more training than individual based surveys do. Because the survey is focused on larger organisations one can also expect that the incidence and amount of training to be higher than in surveys conducted on a more distributed population.

Available from Internet: www.cranet.org [cited 30.7.2004].
Source: Hansson et al. (2004).

Table 4.8. **Studies on human resource management and high performing work systems**

Study, country	Database/Survey	Data	Sample and size	Aim/subject
Laursen and Foss (2003) Denmark	Data extracted from the DISKO project (Database)	Classification of two HRM practices	1 900 firms	HRM practices and their impact on innovation performance
d'Arcimoles (1997) France	French company personnel report	Formal training expenses, level and change data (panel data)	61 firms level data, 42 firms panel data (seven years)	Impact of HRM, wage growth and training on firm performance
Ichniowski et al. (1995) USA	Interviews and company specific data	Proportion employees received off-the-job training (dummy variable)	Longitudinal data on 36 steel production lines, 2 190 monthly observations	Impact of HRM practices on productivity

Source: Hansson et al. (2004).

single variable such as training. However, some studies account for the combined impact of different variables. The study by Laursen and Foss (2003) found training together with performance-related pay as the most important out of nine factors for innovation ⁽⁶⁵⁾. The combined effect of all nine factors proved to be highly significant, which indicates the complementarities of HRM practices.

Also, previous studies came to the conclusion that internally consistent or congruent HRM practices are better predictors of performance than individual HRM practices (Macduffie, 1995; Arthur, 1994). These findings do not mean that training by itself does not have a predictive ability for performance, but only that the combined set of all practices was found to be more informative. The study by Delaney and Huselid (1996) illustrates that the

training measure constantly proved to be significant for performance even in the presence of a large amount of variables capturing other HRM practices.

HPWS (practice) studies are mostly restricted to data at one point in time. Lack of information over a period of time makes the direction of the relationship difficult to analyse. There are some exceptions which measure HRM practices and training over time. However, the results are somewhat contradictory on whether training or HRM generates effects on company performance. The study by Barrett and O'Connell (1999) suggests that training causes productivity effects whereas the introduction of new personnel policies did not show any significant impact on productivity. The results of Bartel (1994) proposed that implementing training programmes generates considerable product-

⁽⁶⁵⁾ These factors are: interdisciplinary work groups, quality circles, systems for collection of employee proposals, planned job rotation, delegation of responsibility, integration of functions, performance related pay, firm-internal training, firm-external training.

Method	Control variables	Outcome measures	Strength/weakness	Findings
Probit model to explain the probability to be an innovator	Sector, firm size, cooperation	Innovative performance	Only innovation performance is the dependent variable	HRM practices do affect the likelihood of a firm being an innovator
Level regression and first difference OLS regressions	Wages, social climate (absenteeism, work accidents, social expenditures), employment variables, etc.	Productivity (change in value added), profitability (return on capital employed)	Impact of change in training measured with lag, HRM controls/weak firm and industry controls	Level of training consistently correlated with level and change in productivity and profitability, change in training associated with change in performance with a two to three year lag (less stable result)
Level and first difference (fixed effect) regression	HRM controls, controls for production line (maintenance, age, raw material, etc.)	Productivity measured by uptime of production line	A number of robustness tests, same production process, panel data/weak training data	Adopting a coherent system of HRM practices produces significant productivity effects. Adopting individual work practice in isolation has no effect on productivity (training)

ivity effects in excess of changes in personnel policies. The study of d’Arcimoles (1997) indicated that training produces substantial effects on both productivity and profitability. A somewhat contradictory view was presented by Ichniowski et al. (1995) where innovative HRM practices have a large effect on productivity while individual employment practices had little or no effect. This result suggests that training by itself is not enough. Table 4.8 provides a summary of selected HRM and HPWS studies with regard to training measures.

Common to these studies is that they found a cause-effect relationship between variables such as training and HRM and performance measures. It is possibly fair to reason that both training and other human resource management practices are important factors in explaining why some firms perform better than others.

3.3.3.2. Training, technological change and innovation

The links between training investments, technological change and their impacts on innovation are rather complex. It is widely agreed that, in the knowledge-based economy, training investment and HRM practices are prerequisites in fostering innovation and realising the productivity potential of new technologies, such as information technologies or advanced manufacturing technologies. To use the potential of new technologies, complementary investments in training and learning are crucial (Laursen and Foss, 2000). There is a growing awareness of the role of internal or organisational factors that mediate the relationship between innovation and firm performance.

According to Hansson et al. (2004), traditional economic-oriented innovation research neglected the internal organisation and human resources of an organisation for a long time. More recent research, however, increasingly pays attention to the role of HRM

and its impact on innovation. An example is Michie and Sheehan (1999) who stress the importance of HRM practices in generating an innovative environment. Their findings suggest that strategies to increase 'employment flexibility' by short-term contracts, weakening trade unions, etc., do not enhance the innovation performance of firms.

Similarly, Baldwin and Johnson (1996) found that more innovative firms place greater emphasis on human resource strategies which foster training. They found statistical evidence that the more innovative firms offered formal and informal training more often, that the training was more often continuous in character, and that these firms more often had innovative compensation packages. An important kind of training in this context, likely to generate high productivity effects, is training associated with the introduction of new technologies or new work practices. This is confirmed, for example, by the case studies carried out by Blandy et al. (2000).

Romijn and Albaladejo (2000) also hold that both training and skills are important determinants of innovative capability. This applies not only to the technical education and prior working experiences of the company owner but also to the technical skills of employees and the amount of training provided. The study by Leiponen (1996a) mentioned above also underlines that innovative firms are more dependent on educational competence in generating profit.

In addition, as Hansson et al. (2004) note, innovation performance is often used as a performance measure for firms. There is broad empirical evidence that innovation is associated with the growth of firms and that in specific industries more innovative firms yield higher financial returns. The results of the Cranet survey (see above) also indicate a strong connection between innovations and a number of variables related to individual workers. The top (10 %) performing firms in their industry sector provided more training, trained more employees, had introduced

supportive HRM policies for employee development to a larger extent and employed more graduates than lower performing firms. This performance indicator represents, in many cases, an important measure for the market value of a firm (see next section).

Finally, research results on the impacts of education and training on companies' productivity and innovations indicate complementarities between different types of education and training (Leiponen, 1996a) as well as complementarities between investments in education/training and in information technology (Gunnarsson et al., 2001). These complementarities produce considerable synergies or externalities. However, the links between the level of education or skills with innovations and productivity is not really surprising as education and skills are generally associated with more complex jobs and increased flexibility.

3.3.4. The market value of a firm ⁽⁶⁶⁾

Research on market valuation is concerned with the factors that determine the growth of individual companies. Studies in this field, and particularly those focusing on education, training, skills and knowledge, are concerned with intangible assets, such as technical knowledge, potential performance and net worth, which are elements of a firm's market value. As Wilson and Briscoe (2004) conclude from their literature review, company performance is closely associated with investments in education and training, as well as skills, which may often be a necessary, if not sufficient, condition to improve these kinds of intangible assets. This research identifies, among other things, R&D, patents and intellectual capital (IC) as key determinants of corporate market value.

3.3.4.1. Indicators for the market value

The (stock) market value of a firm reflects not only its current price but also its expected future performance in terms of company growth, competitiveness and innovative

⁽⁶⁶⁾ This section draws mainly on Wilson and Briscoe (2004) and Hansson et al. (2004).

capability, which is relevant for capital investors ⁽⁶⁷⁾. Although economic and financial indicators still prevail, increasingly indicators of a more qualitative nature are taken into consideration when investing in a company. These are, among others, R&D and patents, trademarks, human capital and investment in human capital. This focus on intangibles is in line with the often propagated shift from a shareholder view, more interested in short-term performance, to a stakeholder view which also takes into consideration the longer-term potential of a company (see for example Westphalen, 2001).

Research studies on the market value of firms are usually based on large-scale panel data sets of individual companies ⁽⁶⁸⁾. However, most research work still focuses on the role of R&D, rather than human capital *per se*. A general result is that the R&D has a significant positive effect on the market value.

Another variable – in addition to human capital, performance and R&D – that has been extensively investigated is the number of patents. This variable, however, will not apply to all sectors, but mainly to manufacturing; the vast majority of companies in the service sector does not have or use patents, for example banking, financing, tourism, etc. Furthermore, the mere number of patents is not very expressive. Therefore, research is emerging on the quality of patents. Here again, research results suggest a strong positive relationship between patents and market value of the companies investigated.

There are close relationships between the variables discussed so far that influence a company's performance in the medium and longer term – patents, R&D, human capital and training. Patents represent an outcome of the R&D process, and R&D depends largely on highly skilled personnel and the acquisition of these skills mainly by education and (continuing) training. Carrying out R&D successfully and obtaining patents requires the deployment of

high level skills although the link between employment of skilled labour, R&D and patents cannot always be proved empirically. New intellectual property variables have recently been tested, including trademark data. Anecdotal evidence suggests that, on average, trademarks are at least as valuable as patents.

Firms with higher training investment per employee yield better profitability and stock market performance in terms of higher stock returns in the following year, higher gross profit margins, higher return on assets, higher price to book ratio, and higher income per employee. This is shown by Bassi et al. (2001, 2002), based on ASTD data (see Box 4.12). The results also indicate that the effects of training emerge with some time lag and that training appears to have long-term effects on profitability.

3.3.4.2. *Training investments as information on market value*

Hansson et al. (2004) propose that investors need more value-relevant information about training investments and their impact on company performance. Investors are often unable to gather this type of information – except some information on labour costs – mainly because of the absence of information on human capital investments in corporate reports ⁽⁶⁹⁾.

However, capital-market actors may also be reluctant to consider research findings on the importance of human capital investments for the following reasons:

- (a) first, there might be a knowledge problem. Capital-market actors might fail to understand the importance of a specific human capital investment and are not aware of available research on the profitability of these investments and the contribution of human capital to the value creation process;
- (b) the second possible reason refers to uncertainty. Even if capital-market actors are

⁽⁶⁷⁾ Note that stock performance may be different from company performance for a number of reasons, for example if stock markets are volatile due to external circumstances (e.g. policy, terrorism, etc.) even if company performance develops well.

⁽⁶⁸⁾ For a literature review see Wilson and Briscoe (2004) and Toivenan et al. (1999).

⁽⁶⁹⁾ If, however, investors had full knowledge of these investments and anticipated the impact on the stock price there would be no effect on the stock performance in the following year.

- aware of the links between human capital indicators and the performance of the firm, they probably do not rely on these indicators in terms of their validity and reliability;
- (c) the third problem refers to ownership. Investor reluctance might be due to the lack of ownership on intangibles such as human capital. Because an organisation does not own individual competence there is a risk of losing this competence;
 - (d) fourth, there could be a management problem in that capital-market actors may not know whether the measures taken by management are relevant and suited to the firm;
 - (e) the fifth and final problem suggested by Hansson et al. concerns the mentality of capital-market actors. They might be neither used nor encouraged to consider human capital investments as important factors that drive firm performance.

To increase the knowledge of capital-market actors and investors about the importance of training investments, valid measurements should be developed and communicated. This refers to activities in the framework of human resource costing and accounting, reporting on human capital and IC ⁽⁷⁰⁾, balanced scorecard, etc. ⁽⁷¹⁾.

Wilson and Briscoe (2004: Section 7.3) point out that 'the management of intellectual capital is still an experimental exercise for most organisations, although a number of key companies [...] are recognised to be leading the process. The Society of Management Accountants of Canada is at the forefront of the promotion of measurement and reporting of intellectual capital. Currently, no company is likely to have a comprehensive intellectual capital management system in place, many will actively manage parts of their intellectual capital, such as their patent portfolios or brands, or parts of their human capital through their HRM systems.'

In this context it is worthwhile to mention the Meritum ⁽⁷²⁾ project under the TSER

programme of the European Commission (Meritum, 2002). The Meritum work was performed from 1998 until 2001 and contained four different activities:

- (a) definitions and classification of concepts, for example intangibles and intellectual capital;
- (b) investigations of how management control of intangibles was performed at firm level;
- (c) capital market implications of the poor information from firms on intangibles;
- (d) development of guidelines for reporting and managing intangibles. The guideline was subject to a Delphi test at the end of the project.

The Meritum work is presently subject to a follow-up project E*Know-net, financed by the European Commission. The aim is to spread the findings from the Meritum work, to improve guidelines and to propose a research and education agenda regarding IC.

As Hansson et al. (2004) point out, the Meritum and E*know-net activities reflect a major transformation in the value creation process of firms whereby intangibles and knowledge increasingly become the major drivers. However, intangible resources cannot easily be identified, measured and reported internally or externally. There is a need to develop a common framework, which involves definitions and classifications of intangibles and a guideline. The main purpose of the E*know-net *Guidelines for managing and reporting on intangibles efforts* is, therefore, to propose new common procedures, documents, rules, etc., which could improve the informative capacity of the firm's financial statements. The guideline provides a set of definitions on intangible resources and intangible activities to be used for the proposed intangible management system (human capital, structural capital and relational capital). Based on the experience of best practice firms, a model for the measurement and management of intangibles is suggested, which

⁽⁷⁰⁾ See for a definition Box 4.3.

⁽⁷¹⁾ Some of these issues have been addressed in the second research report (Westphalen, 2001; Descy and Tessaring, 2001a, Part 3, Chapter 5).

⁽⁷²⁾ *Meritum, measuring and reporting intangibles to understand and improve innovation management*. Available from Internet: www.fek.su.se/home/bic/meritum [cited: 11.08.2004]. For a brief presentation see also the second research report (Descy and Tessaring, 2001a: 217, Boxes 3-26).

covers three different phases: identification, measurement and monitoring of intangibles.

3.4. Discussion

Recent research findings suggest that investments in training generate substantial gains for firms. Furthermore, there is increasing evidence that it is not a matter of whether the nature of training is general or specific, i.e. only useful for the training firm, but more a question of how to stay ahead of competitors. Increasingly firms are financing all types of training, general as well as specific.

Supporting human resource development practices and analysing training needs are seen as important elements in explaining and ensuring the provision of training and training outcomes. Similarly, human resource management practices together with training are associated with firm performance and closely related to a firms' innovative capacity.

The majority of research findings confirm substantial gains for employers from continuing – specific as well as general – vocational training. This is most evident in several studies that connect training investments with changes in productivity, profitability and stock market performance. The majority of these studies also indicate the direction of these relationships, i.e. that training generates performance and not the other way around. Research on firms' investment in human capital and its impact on company performance confirms that the firms' human capital is a major determinant of performance and complementary to technological capital (Ballot, 2003). This applies also to SMEs, although research on the links between training and economic performance of SMEs is not well advanced.

Although comparative international research on human resource management shows that the skills of individual workers are often a critical factor in an organisation's competitive advantage, employers in many countries have a rather negative attitude to investment in skills and training. Others, including capital investors, may underestimate the benefits of training

investment for a company. The finding of Bassi et al. (2001), for example, suggests that most investors do not know about the payoff from investments in training, for example on productivity and stock market performance. Thus the lack of information about training in company reports leads to under-investment in profitable training projects which have a positive net present value.

Unless the institutional and legal infrastructure encourages employers to invest in training, many of them will rely on others to make this investment. Therefore, governments should provide more incentives for companies, for example through tax allowances, direct grants, etc. Such training support can provide strong spill-over effects for society

4. Individual benefits of education and training

The findings presented so far tend to confirm the significant benefits of education, training and skills, or of human capital, for society as a whole and for companies. A plausible assumption is that education and training are also of benefit for individuals, in terms of earnings, career and non-monetary gains.

At microeconomic level, human capital theory views schooling and training as an investment of individuals. Based on rational expectations of returns from this investment, individuals decide to undergo additional education and training if they expect to increase their (lifetime) earnings and if the returns on education and training are higher than returns on alternative investments.

This chapter discusses research on the material and non-material benefits of education and training, or of skills and qualifications, for individuals. The discussion of these benefits takes on two viewpoints: research on the monetary returns on education and training and on individual benefits, including non-material ones, in a life-course perspective. National and cross-national studies and their results are presented ⁽⁷³⁾.

⁽⁷³⁾ Methodological issues have been discussed in Part 2, Section 2.3.

4.1. Monetary returns on investment in human capital ⁽⁷⁴⁾

Research on individual, or private, monetary returns on investment in education and training (or on schooling) are concerned with investigating the links between education/training and performance at individual level. They carry direct implications for the macroeconomic level discussed earlier in this part which can, more or less, be regarded as the aggregation of the individual returns. Although most of research studies emphasise the returns on higher education, there is a fairly extensive literature on the returns on vocational training.

However, there is considerable variation in individual returns between European countries and many results are hardly comparable due to a lack of sufficient cross-national surveys and differences between data and estimation methods. Similar problems arise when comparing studies for one country, even if based on the same data sets: their results often differ considerably between studies due to different model specifications, inclusion of control variables, etc.

The effect of formal schooling on wages has to take into account that individuals with high and low skills do not only differ in terms of the time they spent in education and training, but also in other characteristics such as ability or family background. However, some of these characteristics – and in particular ability – are difficult to observe. If ability is positively correlated with (successful) participation in further education and training and is not controlled for, the mere effect of education and training would be overestimated: more able individuals tend to participate more in advanced education and training. A study that does not control for this factor is likely to be

affected by an ‘ability bias’, or more broadly a ‘selectivity bias’ ⁽⁷⁵⁾.

Rates of return (RoR) on education or schooling calculate the return on the resources invested in education, taking into account the costs. RoR denote the marginal internal RoR of an additional education and training (or year of schooling) compared with the next lower level of education and training. De la Fuente (2003: 2) defines the private RoR on schooling as follows: ‘the rate of return to investment in education is the ratio of the expected annual increase in net earnings due to an additional year of schooling divided by its cost, and adjusted for the finite lifetime of the investment (the working life of the individual).’ RoR can be calculated either by a discounted cash flow or applying the Mincerian wage equation (see Part 2, Section 2.3.1).

Two RoR are distinguished:

- (a) private RoR, considering private benefits and costs (opportunity costs and direct costs);
- (b) social RoR, which include the public costs of education and tax returns, but do not account for social benefits such as externalities.

Ideally, the social RoR would relate all resources invested to all benefits of education.

4.1.1. Private returns on education

Some recent results of research into the private returns on education (or schooling) are presented below.

Ashenfelter and Rouse (forthcoming), based on data for identical twins in the US, find that for each year of additional schooling, the twin with more education earned 10 % more per year of additional schooling than his or her identical twin. This result was obtained after correcting for measurement error or quality of schooling. This implies, for example, that higher education graduates who need four additional years to complete a degree should

⁽⁷⁴⁾ This section draws mainly on Wilson and Briscoe (2004). The authors refer in particular to recent reviews on the rates of return on education and training undertaken by Harmon and Walker (2001) and Blundell et al. (2001). A review of the most significant literature on rates of return on continuous vocational training in companies has been carried out by Barrett et al. (1998) in Cedefop’s first research report (Tessaring, 1998). Therefore, we refer to the first research report and will present in this section only some updates and further discussions.

⁽⁷⁵⁾ The approaches to resolve the difficulties caused by unobserved characteristics are discussed in Part 2, Section 2.3.1.2.

Box 4.15. Public funding and private returns on education (PURE)

The overarching objective of PURE is to study the impact of different systems of public financial support for school attendance and of differences in educational differentiation and school admission rules (free or selective entry) on observed outcomes in the labour market, in particular, in terms of the levels and dispersion of private returns on education and education-related inequality in earnings.

Detailed objectives of the study:

- (a) analysis and comparison of wage and human capital structures and private returns on education between and within countries over time to uncover distinct trends as well as similarities and dissimilarities across countries;
- (b) analysis of the impact of country-specific trends in education returns over time and underlying market forces (supply-side and demand-side factors);
- (c) analysis of carefully differentiated measures of private returns by type and level of education in order to highlight and compare national systems of education;
- (d) analysis of the structure and evolution of national education systems, admission rules and systems of financial support for school attendance to be used as an input in (e);
- (e) analysis of the effects in differing systems of public funding and admission rules on private returns on education and on earnings inequality related to differences in educational attainment.

The researchers examine broad national data sets containing extensive individual-level information on wages and education (based on repeated cross-section and some longitudinal data), and link the observed patterns and trends to national educational systems and policies. Depending on the country considered, data covered the past 10 to 25 years with the last observation year in the mid-1990s.

The project involves 15 European countries: Austria, Denmark, Finland, France, Germany, Greece, Ireland, Italy, the Netherlands, Norway, Portugal, Spain, Sweden, Switzerland and the UK.

Results

- (a) for most countries, the value of RoR to education lies between 8-10 % (average: 9.7 %). In Sweden it is less than 6 %, possibly because of wage compression; in Portugal the RoR is more than 12 %, and in the UK around 14 %;
- (b) the RoR (in the mid-1990s) seem to fall into three different classes: the lowest return (4-6 %) on one additional year of education is found in the Scandinavian countries (Denmark, Norway, Sweden) and in the Netherlands; the highest RoR (around 10 %) are found in Ireland and the UK; in the other countries the RoR lie in-between;
- (c) in some countries (Germany, Greece, Italy, Ireland and the UK) there is a substantial variation in RoR between gender, i.e. returns for women are significantly higher than for men;
- (d) countries that experienced a downward trend in educational returns during the past decades were Austria, Sweden and Switzerland; countries with an upward trend are Denmark, Finland, Italy and Portugal. For the other countries no clear pattern of trends could be found;
- (e) net after-tax RoR are in most cases much higher than before-tax real returns on alternative assets (e.g. bonds, equity);
- (f) public policies have an important effect on the private RoR. Government subsidies to education are large (from 20 % to more than 60 %) and more than compensate for disincentives generated by progressive taxes and social benefits;
- (g) high student unemployment rates reduce the RoR, in particular in Italy, southern EU and Scandinavia;
- (h) average RoR exceed average returns to government debt and corporate equity; this means that education and training is an attractive investment alternative from the individual perspective.

Sources: available from Internet: www.etla.fi/PURE [cited: 11.08.2004]; PURE (Final report, 2001); Harmon et al. (2001).

earn about 40 % more than grammar school graduates (or a similar level).;

Harmon et al. (2001) carried out a meta-analysis of a large number of wage equation ⁽⁷⁶⁾ estimates undertaken or collected as part of PURE, a large research project on the returns on education in Europe funded by the European Commission (Box 4.15). They find for 15 European countries (mostly EU Member States) that the average return in Europe is 6.5 % and that country means range

from around 4.5 % to 10 %, with some Scandinavian countries at the lower end of the distribution (around 4 %) and the UK and Ireland at the top (around 12 %).

Similar evidence is given by Denny et al. (2001) for different European countries. They find large differences, Norway with the lowest, Ireland and the UK with the highest returns. These authors also observed a decline in returns on schooling in Europe in the 1960s to the mid 1980s. Since then, until 1997 (last

⁽⁷⁶⁾ Based on the Mincerian earnings equation (Part 2, Box 2.12).

Table 4.9. **Marginal RoR to education (a) in Europe, 1995**

Country	Medium versus low level of education (b)	High versus medium level of education (b)
Sweden	5.8	6.3
Finland	6.2	7.6
Italy	6.9	6.4
Spain	9.3	8.6
Switzerland	9.3	8.9
France	10.3	8.0
Germany	14.4	6.1
Austria	15.0	5.7
Ireland	15.9	11.5

(a) adjusted for employment probability and expected unemployment benefits;

(b) low: ISCED 0-2; medium: ISCED 3; high: ISCED 5-7.

Countries sorted by comparison between medium and low level.

Source: Barceinas-Paredes et al. (2000: 19).

observation year), returns tended to increase, in the US even more than in Europe.

Also in the framework of PURE, Barceinas-Paredes et al. (2000) have calculated adjusted (77) marginal RoR for 1995 using the Mincerian earnings equation to predict the expected life-cycle earnings flow. 'Marginal' in this context means the comparison of a given educational level with the respective lower level, i.e. the additional monetary benefit people could expect when deciding on a higher level of education and training. Table 4.9 shows the results.

The results show— with the exception of Sweden and Finland – higher RoR for people with upper secondary education compared with lower educational levels than is the case for higher compared with medium educational levels. Particularly in Germany and Austria, these differences are considerable (78).

Concerning wage differentials between women and men, the same authors find (in the framework of PURE, see above) that RoR on

schooling are greater for women than for men in Europe. The effect of schooling on female RoR exceeds male RoR by 5 percentage points in Ireland and by 2 or more percentage points in Italy, West-Germany, Greece and the UK. Denny et al (2001) observe that the differential is greater in countries with lower female labour force participation.

Table 4.10 based on a recent OECD study (2003b) shows the marginal private and social (79) RoR on upper secondary and higher education, by gender. The results suggest that:

- private RoR on upper secondary compared with lower secondary education for males range between over 6 % and over 16 % in OECD countries; Sweden, Japan and the Netherlands have the lowest, and the US, the UK and France, the highest RoR. Female RoR range between almost 7 % (Germany) and 19 % (France);
- private RoR on tertiary compared with upper secondary education for men range between 6.5 % and more than 17 %; the lowest are observed for Italy, Japan and Canada, and the highest for the UK and the US;
- private RoR in general are around 10 % and higher except in Japan, Sweden and the Netherlands for upper vs lower secondary education, and in Canada, Germany, Italy and Japan for higher vs upper secondary education.
- private RoR for upper secondary compared with lower secondary education are higher for men than for women, except in France, Japan and the Netherlands; for higher education compared with upper secondary, the pattern is similar with the exception of France where the rates for men are higher than for women.

De la Fuente and Ciccone (2002) note that wage equation coefficients have to be treated with some caution, for example because of the influence of labour-market institutions, social norms and different wage-setting practices.

(77) The adjusted internal RoR takes into account the employment probability and the expected unemployment benefit. It therefore captures unemployment risks (variation of the employment probability) as well as institutional factors (unemployment benefit).

(78) Some more results of PURE, using longitudinal data, are presented in Box 4.15.

(79) Social rates of return will be discussed in the next section.

Table 4.10. **Private and social RoR on education in selected OECD countries by gender 1999-2000 (in percentage points)**

	RoR on upper secondary compared to lower secondary education ^(a)				RoR on tertiary education compared to upper secondary education ^(a)			
	Males		Females		Males		Females	
	Private RoR	Social RoR	Private RoR	Social RoR	Private RoR	Social RoR	Private RoR	Social RoR
Canada	13.6	m	12.7	m	8.1	6.8	9.4	7.9
Denmark	11.3	9.3	10.5	8.7	13.9	6.3	10.1	4.2
France	14.8	9.6	19.2	10.6	12.2	13.2	11.7	13.1
Germany	10.8	10.2	6.9	6.0	9.0	6.5	8.3	6.9
Italy ^(b)	11.2	8.4	m	m	6.5	7.0	m	m
Japan	6.4	5.0	8.5	6.4	7.5	6.7	6.7	5.7
Netherlands ^(c)	7.9	6.2	8.4	7.8	12.0	10.0	12.3	6.3
Sweden ^(d)	6.4	5.2	m	m	11.4	7.5	10.8	5.7
United Kingdom	15.1	12.9	m	m	17.3	15.2	15.2	13.6
United States	16.4	13.2	11.8	9.6	14.9	13.7	14.7	12.3

^(a) comprehensive rates of return;

^(b) data for males derive from 1998 post-tax earnings data;

^(c) 1994;

^(d) for women, earnings differential between upper secondary and lower secondary levels are not large enough to permit a positive RoR calculation.

m = missing data.

Source: OECD (2003b: Tables A14.3 and A14.4); see this source also for methodological explanations.

Making allowance for some of these distortions, they assume that the 'true' return on schooling ⁽⁸⁰⁾ lies somewhere between the central estimate of 6.5 % found by Harmon et al. (2001) and their average estimate of 9 % for the UK and Ireland, countries that appear to have the most flexible labour markets in Europe.

It should be noted that RoR on schooling usually do not consider, among other things, non-material returns and the impact of education and training on labour force participation and employment probabilities. As a result, they tend to underestimate the 'true' returns on schooling by an amount that may be large but which is extremely difficult to measure with precision (de la Fuente and Ciccone, 2002).

4.1.2. Private vs social RoR

An important feature of research into the RoR on education and training is the distinction between private and social RoR (Wilson and Briscoe, 2004).

Whereas private RoR relate only to the costs and benefits of the individual undertaking the education or training, the social RoR takes into account all costs and benefits, i.e. including those paid for or accrued by society, for example returns of income taxes. Private and social returns are generally different due to the fact that, for example, considerable costs are born by the government (subsidies for schools, training providers, higher education) but are not returned to the same extent to society (e.g. by taxes) in terms of monetary benefits.

However, the social RoR do not include social benefits in excess of private benefits of education (see for example Arias and McMahon, 2001). In addition, social RoR do not account for spill-over effects and externalities from education and training ⁽⁸¹⁾.

Social RoR are generally somewhat lower than the private ones (see the OECD analysis in Table 4.10). In European countries, the social RoR on tertiary education for men is

⁽⁸⁰⁾ Using the Mincerian wage equation; see Part 2, Section 2.3.1.1.

⁽⁸¹⁾ Several examples and research findings for these effects, including non-material benefits have been discussed throughout in this part, see for example Section 2.1.2.

around two percentage points lower than the private return; the outliers are Denmark and Sweden where it is four or more percentage points below. For women, the gap between the social and the private return is similar, with again the former outliers and the Netherlands where the private return exceeded the social return by five to six percentage points. Social and private RoR of upper-secondary education follow a similar pattern.

Somewhat surprising is the finding that social RoR – for tertiary vs upper secondary education – are higher than the respective private RoR in France and Italy (men). This could be due to higher taxes on earnings or lower costs of tertiary education compared with upper secondary education.

Mingat and Tan (1996) estimated the social returns on education, by incorporating externalities and non-economic effects, for a wide range of countries over the period 1960-85. Their results confirm the social profitability of investing in education, but they also suggest that the results are sensitive to the stage of economic development achieved by the individual country.

4.2. Benefits of education and training in the individual life course ⁽⁸²⁾

The previous section reviewed studies which try to calculate RoR, using data for one single point in time, by estimating lifetime earnings ⁽⁸³⁾. However, earnings differentials may not remain constant over time. This would be the case if individuals – perhaps also encouraged by information on favourable RoR in the past – increasingly decided to undergo further education and training and labour-market demand did not increase to the same extent as the supply of more highly skilled

people. In such a supply surplus, earnings and RoR for highly skilled people would decrease and those for the lower skilled increase ⁽⁸⁴⁾.

A more sophisticated (and demanding) line of research is to calculate the benefits and costs of education and training over the life course of individuals by using panel or longitudinal data.

As discussed in Part 2, Section 2.3.2, research focusing on the lifetime follows two main strands:

- (a) life course research which concentrates on the institutionally offered situation, well-defined alternatives (mostly guaranteed by law) and subsequent decisions and pathways chosen by individuals;
- (b) biographical research that focuses on the individual subjective perception, assimilation, understanding, and reconstruction of reality (Heise and Meyer, 2004: Chapter 3, with reference to Meulemann, 1990).

Life course and biographical research were not primarily developed to investigate the benefits of education or training, but for research on social and demographic change processes. Originating from national social research movements in the US and northern and central European countries (amongst them France, Germany, Norway and the UK) empirical life course research has developed considerably in the past two decades. Increasingly it also looks at internationally comparative issues and further development of methods and databases. Current subjects of investigation are societal trends, comparisons of different political systems and national differences in the access to public goods, such as education and training ⁽⁸⁵⁾.

Life course and biographical research became a specific scientific research field also in education and training research. It is embedded in major debates on education and training issues of the past decades, in particular: educational expansion, equal

⁽⁸²⁾ This section draws mainly on the compilation of research studies by Heise and Meyer (2004).

⁽⁸³⁾ This can be done, for example, in a Mincerian earnings equation based on data for a number of individuals, by calculating age-dependent earnings and assuming that these reflect the lifetime earning stream.

⁽⁸⁴⁾ In the neoclassical view, this would lead to a subsequent adaptation of supply with more people not choosing higher education, but this takes time.

⁽⁸⁵⁾ Above, we have argued, however, that education and in particular training are no pure but semi-public goods (Section 2.1.1).

access to education and training and equal opportunities, social selection and ability, and individual performance in terms of literacy and numeracy (see for example the OECD's IALS and PISA studies; OECD 2001a, 2003a; OECD and Statistics Canada, 1997, 2000). The principle of free individual decisions within education and the right to follow individual aspirations made research on the factors of influence on these decisions, and on the impacts for individuals and societal processes, highly relevant.

Using longitudinal surveys or panels and looking at the pathways of different generations, or 'cohorts' (see Box 2.13 in Part 2), through education, training and employment is a precondition for answering a number of questions. For example, the question of equal opportunities cannot be analysed at one single point of time in the educational career. Rather, the circumstances at the beginning of the process and the process itself are important determinants to be considered (Meulemann, 1990). Equally, individual aspirations, abilities and performances are not stable over time; they change with personal and societal development.

In European social research, exploration of educational benefits in the life course has not yet received full attention. While work is one of the central themes in sociology and economics, and labour-market processes received great scientific attention since the labour-market crisis in Western Europe in the 1970s, the benefits of education and training for the occupational life course are still marginal topics in this field. Moreover, research activities in countries with a long tradition in life course analysis seem to shift emphasis away from researching the impact of education to broader investigations on individual life chances under certain social conditions and developments, where education and training play a role *inter alia*. Reasons may be difficulties regarding definition and measurement of education, training, skills, abilities and qualifications which make hypotheses on educational outcomes less attractive than the mere description of social situations and

developments. Therefore, as Heise and Meyer (2004) note, research on educational and occupational careers is limited compared with the large body of research on education and human capital.

Some topics are of particular relevance for life course and biographical research in education and training because of their time-dependent character.

- (a) Transition from (compulsory) school to training and higher education. The decision to continue education and training after compulsory school (or to leave education) is influenced by individual aspirations and ability but also by the provision of further education and training places, expectations concerning future career and the attractiveness of different training courses. Social background, vocational guidance and counselling, information on future perspectives and supportive measures are further parameters that influence these decisions; in many cases a choice made at early age will more or less determine the future life course of an individual;
- (b) Labour-market entry. Increasing unemployment in Europe, and particularly high and persisting youth unemployment, gave rise to research on the circumstances and conditions to improve the entry or return of unemployed people into the labour market and the transition from education to employment (e.g. Farvaque and Salais, 2002; Jahnukainen, 2001; Kortteinen and Tuomikoski, 1998). Following the careers of people in different situations and investigating the measures and programmes to improve these transitions has been one major issue of life course and biographical research and requires the use of longitudinal data;
- (c) Career mobility and lifelong learning. In view of accelerating social, economic and technological change, adequate skills and lifelong learning are increasingly becoming prerequisites for getting and keeping a qualified job. This includes transitions into initial and continuing training – formal or non-formal – in the course of life, followed

Table 4.11. **Selected studies on benefits of education and training from a life course perspective**

Country	Study	Data source ^(a)	Outcome variable	Explanatory variable	Benefits ^(b)
National studies					
DK	Skov (1998)		educational career	level of education or qualification; VET	material
FIN	Nummenmaa (1996)	OSLM	educational career, occupational career, personal development	level of education or qualification; VET	material; non-material (+/0)
FIN	Koivuluhta (1999)		educational career, occupational career	level of education or qualification; VET; type of qualification (subject)	material; non-material
D	Pannenberg (1995)	GSOEP	occupational career, income development	further VET	material
D	Becker and Schömann (1999)	GSOEP; GLHS	income development	further VET	material (+)
D	Diewald and Sørensen (1996)	GLHS	occupational career	level of education or qualification	material (+)
D	Blossfeld and Timm (1997)	GSOEP	marriage	level of education or qualification; years of education/training; family background	non-material
D	Becker (1998)	GSOEP; GLHS	life expectancy	years of education/training	non-material (+)
D	Becker and Schömann (1996)	GLHS	income development	further VET	material (+/-)
D	Born (2000)	SFB 186	occupational career	type of qualification (subject)	material
D	Bender and Dietrich (2001)	social insurance data	occupational career, unemployment risk	level of education or qualification	material (+)
D	Hillmert (2002)	GLHS	occupational career	level of education or qualification; type of qualification (subject)	material (+/0)
IRL	Keogh and Downes (1998)	individual autobiographical summaries	vocational training opportunities	further VET	material; non-material
IRL	Gash and O'Connell (2000)		educational career, occupational career, income development	level of education or qualification	material (+)
NL	Wolbers (2000)	OSA	unemployment risk	level of education or qualification	material (+)
CH	Schallberger et al. (2001)	ZLSE	occupational career, personal development	level of education or qualification; VET; type of qualification (subject)	material; non-material
CH	Meyer (forthcoming)	TREE; PISA	educational career, occupational career	level of education or qualification	material
UK	Payne (1995)	YCS	educational career	level of education or qualification; type of qualification (subject)	material
UK	Dale and Egerton (1997)	NCDS	occupational career, family formation, income development	level of education or qualification; type of qualification (subject); family background	material (+/-)
UK	Bynner, et al. (2001)	NCDS; BCS70	income development, occupational career, health	skills	material; non-material

Country	Study	Data source ^(a)	Outcome variable	Explanatory variable	Benefits ^(b)
UK	Bynner and Parsons (2001)	NCDS; BCS70	unemployment risk	years of education/training; skills	material; non-material
UK	Unwin and Wellington (2001)	MA	learning, skills, transitions from school to work	VET; type of qualification (subject); skills	material; non-material (subjective)
Cross-national studies					
A, DK, FIN, F, D, I, P, UK, NL, B, IRL, EL, E	Brunello (2001b)	ECHP	VET, RoR	years of education or training	material (+)
A, CH, DK, FIN, F, D, I, NO, P, UK, NL	Brunello and Comi (2000)	11 national data sets (GSOEP, OSA, LFS, etc.)	experience, RoR	years of education or training	material (+)
I, F, E, FIN, UK, D, S	Kieselbach (2000)	qualitative interviews; official statistics	non	general education; VET; further VET	material; non-material
UK, D	Hillmert (2001)	BHPS, GLHS	occupational career	general education; VET	material
D, PL	Schömann (1994)	GLHS, PSSM	RoR	general education; VET	material (+)

^(a) See Frequently used abbreviations at the beginning of this publication.
^(b) and result (positive/negative) if indicated in the study.

Source: Heise and Meyer (2004).

by periods of employment or non-employment. In this context, flexibility of the labour market is seen as an important precondition to reduce unemployment considerably and to create new jobs. However, this also implies that individuals might change their job several times in their lifetime, with interrupting phases of unemployment or further training. According to the individualisation thesis of Beck (1994), ‘patchwork’ careers will increase and have substantial societal, socio-psychological and economic impacts ⁽⁸⁶⁾. The use of longitudinal data and event history analyses are adequate ways to investigate these processes.

The significance of life course and biographical approaches also becomes clear from research findings that time spent inside and

outside employment not only depends on specific exogenous situations such as business cycles and labour-market imbalances but also on individual experiences and skills acquired in the life course. Both effects can be separated and controlled for when analysing cohort processes in order to test the ‘patchwork’ career hypothesis as a phenomenon of structural change.

4.2.1. Research findings at national and European level

Table 4.11 provides an overview of selected studies touching on the benefits of education and training. The results of these and other studies will be commented upon below. However, we should emphasise that life course studies on individual benefits of education and training carried out at national level are rather diverse and, therefore, can hardly be

⁽⁸⁶⁾ For empirical research work on the risks of this kind of job career see for example Andreß (1989); Büchel (1992); Felstead et al. (1997); Tuominen (2000); van de Werfhorst (2002).

compared. Reasons are different definitions of the explanatory and dependent variables and the often diverse research designs, research disciplines and traditions in different countries.

Research studies on individual work histories over the life course normally delimit the explanatory variables 'education' or 'training' to the level and/or years of education or training. Furthermore, with the exception of few studies (e.g. Born, 2000; Dale and Egerton, 1997) education and training are treated as explanatory variables among others such as work experience, family background and gender to explain differences in the economic, social and private well-being of individuals. This denotes a considerable gap in life course research, because too little can be said about differences by type and quality of education and training, and of qualifications and skills.

Furthermore, specific educational benefits are hardly determinable by quantitative methods, because causes and effects are interrelated. For these reasons, the list of existing research work on individual life course benefits of education, training and skills includes several empirical works on education and training benefits in a lifetime perspective which originate not primarily from life course research but use longitudinal data. Our review of research work addresses the following issues relating to the benefits of education, training and skills:

- (a) monetary returns and lifetime income;
- (b) impact on labour-market participation;
- (c) education, training and transitions;
- (d) generational and cohort differences;
- (e) social differences;
- (f) non-material benefits and subjective biographical perception.

The theoretical background for selecting these issues has been discussed in Part 2, Section 2.3.2.1.

4.2.1.1. *Monetary returns and lifetime income*

There are few 'genuine' life course studies, based on panels or longitudinal surveys, which analyse lifetime monetary returns on education

(in particular earnings). The results of some selected studies are presented below:

- (a) Steiner and Lauer (2000) examined the returns on education in Germany from 1984 to 1997 on the basis of the German Socioeconomic Panel (GSOEP) and showed that rates of return are not constant over the life cycle. Women experienced higher RoR (10 %) than men (8 %). However, female RoR declined after the mid-1990s, in particular for younger women. The authors assume that this reduction was a result of educational expansion in Germany since the 1970s which favoured women and led to an disproportionate increase in the supply of highly skilled people;
- (b) Becker and Schömann (1996, 1999) estimated the long-term educational returns on further VET based on the German life history study (GLHS) and the GSOEP. Their results suggest a general tendency to lowering RoR on further VET in Germany. A reason could be that individual training costs (including opportunity costs) have increased and overcompensated, in a RoR approach, the earnings increase;
- (c) Barceinas-Paredes et al. (2001) analysed longitudinal data for France and Spain and asked whether seniority has an effect on income over time. In France, educational returns start to decline after five years of job tenure, whereas in Spain, returns do not decrease before 20 years seniority. Unfortunately the authors of the article do not give any information on the data sets used for their analysis.

Many studies, especially those concerned with RoR on education in the framework of human capital theory (see Section 4.1) do not have data which allow to apply a 'genuine' life course approach. Instead, they enhance their approach – which is mostly based on cross-section data derived from surveys – by calculating lifetime earnings ⁽⁸⁷⁾. Despite some reservations (see above), this approach – even

⁽⁸⁷⁾ For example, by reinterpreting data on 'age-earnings profiles' as 'time-earnings profiles', or by using the Mincerian earnings equation (see above) and thus assuming that the present earnings of older people represent the future earnings of today's younger generations.

if it is not a real life course approach – contributes to better understanding of the dynamic impacts of education and training.

These studies assume a direct relationship between the accumulation of education, training and skills and the rise of current and lifetime earnings (Brunello and Miniaci, 1999; Goux and Maurin, 1994). Several studies confirm both the relevance of skills and qualifications acquired through initial and continuing education and training for individual earnings at a certain point of time as well as higher rates of monetary returns in a lifetime perspective.

Education and training affect earnings not only *per se* but also, in a complementary or cumulative way, because participation in continuing training is higher among individuals with more education (Brunello, 2001a; see also Section 4.2.1.3 below). However, these cumulative effects – and in particular the effects of training – seem to weaken after a few years and even turn negative if no investment in further training is made.

Therefore, education and training have only temporary impacts on monetary returns. RoR on education and training are not constant over the life cycle. While returns may rise in the first years of the working career, the increase may lower in mid working life, stagnate and finally may decline in the second half until retirement.

Cross-country comparative studies from a life course perspective are still rare in Europe and those existing are limited in their explanatory power and comparability, mostly due to differences in the national surveys they are based on. Existing comparative studies on educational returns and individual working careers show a broad range of educational benefits across European countries. According to studies by Harmon et al. (2001) and Asplund and Pereira (1999) people in Scandinavian countries have relatively low rates of returns on education and training, whereas people in Ireland and the UK show

high rates over time. People in central European countries are located somewhere between these two extremes; some show a downward trend (such as in Germany and Austria), whereas others (such as in Portugal or Italy) remain stable or even display slight upward trends (Chevalier, 1999; Brunello and Miniaci, 1999; Goux and Maurin, 1994).

Other examples are Brunello and Comi (2000) and Brunello et al. (2001a) who analyse the relationship between education and earnings growth in 11 countries. They find evidence in all countries that education and training provide not only an initial job entry advantage but also a permanent advantage that increases with time in the labour market. Analysing this relationship in more detail they show that differences in earnings growth associated with education are higher in countries which have experienced both relatively fast labour productivity growth and relatively low educational attainment (e.g. Italy and Portugal). Equally, earnings growth differentials are higher in countries with more market-led education and training systems and lower in countries with a higher level of corporatism⁽⁸⁸⁾, such as Denmark, Germany and Norway⁽⁸⁹⁾.

The project PURE (Public funding and private returns to education; see Box 4.15) examined national educational returns of 15 European Member States not only in regard to wages but also to better job opportunities. The results suggest clear national differences in job status and trends of return rates. Although the principal research perspective was not a life course approach, results from (partly) longitudinal measurement point to several life course relevant aspects.

The explanatory power of RoR on education and training in the life course, based on human capital theory, is still limited although substantial progress has been made over recent years. Refinements of estimation techniques made it possible to improve analyses of the effects of education, training and skills.

⁽⁸⁸⁾ 'Corporatism' in this context refers to an education and/or training system which is regulated, based on consensus, by representatives of various social groups (e.g. social partners, government, local bodies, etc.).

⁽⁸⁹⁾ This tendency is also confirmed by the RoR calculated by OECD (see Table 4.10).

4.2.1.2. *The effects of education and training on labour-market participation*

Another group of studies investigate the links between education and training, and individual occupational careers or participation in the labour market. Some studies are rather descriptive, others are based on specific labour-market theories. However, the diversity of subjects investigated, of theoretical approaches and methods hampers comparison of results, in particular from a European comparative perspective.

Most studies confirm that higher levels of education and training reduce the probability of unemployment and increase labour force participation ⁽⁹⁰⁾. This is true for most – but not all – European countries. The differences between countries are sometimes large. Indicators presented in Part 3, Section 1.5 confirm this link.

The effect of education and training is twofold: basic education and training determine the access to higher and continuing education and training to a considerable extent, and individuals with higher educational and vocational degrees have better further training and career opportunities. It can be assumed that this selective mechanism of participation in further training tends to lead to a cumulative increase in educational inequality between high and low skilled people over the life course ⁽⁹¹⁾.

Again, there are considerable differences across countries regarding the link between education, training and career prospects over the life course. In countries with highly standardised training systems, characterised by well defined vocational qualifications such as in Germany's dual system or in countries with national qualification standards, benefits of education and training in terms of employability, job stability, avoidance of unemployment and successful working careers (and associated higher earning) tend to be generally higher (Fitzenberger and Prey,

1999; Shavit and Müller, 1998) than in countries with less integrated education and employment systems and lower vocational specialisation (Liefbroer, 2002; Klijzing, 2000; Blossfeld and Mayer, 1998). However, in systems with high standards of vocational qualification, but low permeability between education and training routes, the gap between the low qualified and higher qualified continually rises over the working life and the potential for compensation for disadvantages experienced in the early career is rather limited (Antoninis and Tsakoglou, 2001).

A detailed analysis of the impact of on-the-job and off-the-job training on the working career is provided by Pannenberg (1995). Besides the measurement of income development he looks at individual job mobility as a possible benefit of education and training programmes taking place after compulsory schooling, mostly in companies. He finds that short (two to seven days) and longer (one week to one month) training on-the-job increases the possibility of upward mobility significantly. Shorter training periods tend to have more positive impact on in-company mobility whereas longer training programmes increase inter-company mobility. Off-the-job training for unemployed people significantly reduces the risk of staying unemployed only if programmes are medium-term (6-12 months). However, longer term programmes (more than 12 months) increase the risk of staying unemployed.

At European level, projects funded by the European Commission that investigate educational benefits in the life course are FAME (Vocational identity, flexibility and mobility in the European labour market), the TSER projects Newskills (New job skill needs and the low skilled) and Yuseder (Youth unemployment and social exclusion: dimensions, subjective experiences and institutional responses in six countries of the EU) ⁽⁹²⁾, the Leonardo da Vinci project

⁽⁹⁰⁾ For example Heinrich and Hildebrandt (2001); Bukodi and Robert (2002); Noguera et al. (2002); Becker and Schömann (1999).

⁽⁹¹⁾ For example Solga (2002); Stone et al. (2000); Bynner (1995); Dietz and Matt (1994).

⁽⁹²⁾ The project was carried out between 1998 and 2000 within the TSER programme of the European Commission. Available from Internet: www.ipg.uni-bremen.de/research/yuseder [cited 19.7.2004].

Duoqual (Dual qualifications and vocational mobility) and VTLMT (Education, vocational training and labour-market transitions). These projects have already been presented in Descy and Tessaring (2001a).

Even considering national differences, the material benefits of education and training in terms of higher labour force participation, higher entry wages, higher wages throughout the working life, increased job opportunities and a reduced risk of becoming unemployed are virtually universal (Heise and Meyer, 2004).

4.2.1.3. *Education, training and transitions*

Transition and mobility research is booming in many European countries. This can presumably be attributed to the aggravation of the labour-market situation, in particular for young school leavers, graduates and the unemployed. Transitions between education and training, education/training and employment, etc., are of special interest for life course research because they are sensitive phases that are highly relevant for the later working career of an individual. Life course progression is made up of successive transitions.

The successful transition from education or training to the first job is formative for the whole later career and life of an individual. Schaeper et al. (2000) who investigated working careers up to eight years after the transition from VET to work in Germany found that steady working careers depend directly on the first occupation trained for and indirectly on the educational background of an individual. A study by Bender and Dietrich (2001) for Germany confirms the long-term consequences of successful or less successful transitions from education and training to work. National longitudinal data sets have been used widely to study the processes of transitions of young people from education/training to working life (see also and for details the first and second research reports of Cedefop [Tessaring, 1998; Descy and Tessaring, 2001b]). Comparative research indicates considerable differences across

countries in the job entry of education and training leavers ⁽⁹³⁾.

Transition patterns vary considerably across countries (Starrin et al., 2000). They seem to be easier in Denmark, Germany, the Netherlands and Austria than in Southern European countries like Spain, Greece or Italy where direct entry into a stable position after completing education or training seems to be extremely difficult. In Belgium and Sweden, where the systems of education and training are decoupled with an emphasis on general education, a trend of polarisation between high and low skilled people can be observed. Lack of work experience and job related skills makes it more and more difficult to gain a first job (Rantakeisu et al., 2000). This is also true for Germany and its dual system of initial training, which traditionally has been known to facilitate transitions from training to work. Due to a deterioration of the labour-market situation and cutbacks in jobs in recent years, the dual system experienced a serious 'crisis'. Companies reduced their training offers or did not take on their apprentices. This situation has worsened recently and has led the German government in 2004 to introduce a training levy for non-training firms.

In southern European countries like Spain and Greece, weak links between education and the labour market are responsible for many of the problems regarding youth unemployment. In both countries youth unemployment is three times higher than the overall unemployment rate (Part 3, Section 1.7). Despite poor prospects, university degrees in Greece are overrated, because they are traditionally connected with high social esteem, and minor value is attached to vocational training. Spain is currently trying to reduce its high youth unemployment by system reform aiming to improve the adjustment of demand and supply. The decision either to go to university or to choose a vocational strand is still mainly determined by financial opportunities and social class. However, people – in particular males – with a higher education degree are

⁽⁹³⁾ For example, Liefbroer (2002); Gash and O'Connell (2000); Fabrizio (2000); Koivuluhta (1999); Bratberg and Nielsen (1998); Breen (1995); Arum and Shavit (1995); Bynner (1995).

Box 4.16. Life courses in the globalisation process (Globalife)

The Globalife project is a multidisciplinary and international comparative research programme funded by the German Volkswagen Foundation from 1999-2004. It is centred in the Faculty of Sociology at the University of Bamberg.

The project goal is to study the influence of globalisation on life courses in OECD-type societies.

Globalisation is certainly not a coherent process and cannot be described as a single phenomenon. In socioeconomic terms, it is a series of significant macro processes that are common to all modern societies:

- the increasing significance of knowledge and information;
- the extraordinary rise in productivity;
- the growing need for flexibility;
- the rising uncertainty of future developments;
- intensified competition among individuals, firms and nation states.

These macro processes are closely linked with the rapid diffusion of new information and communication technologies, which reduce the importance of spatial distances for all kinds of interactions and make changes in economic and social life faster and less predictable everywhere.

Individuals, however, have to make long-term binding commitments at various phases of their life courses. For example, they have to opt for educational tracks, specific jobs, career paths, particular partners, having a child or not; these are decisions which, once taken, are hard to revise. In other words, in modern societies there is an increasing tension between growing uncertainty of future life course situations and individuals' natural needs to make self-binding decisions. The central question for the project is, therefore, how actors deal with this new kind of life course situation in various societies. A major hypothesis at micro level is that greater uncertainty forces actors to simplify

action situations to less complex patterns.

The major question of the project is, to what extent do modern life courses develop in nation-specific, path-dependent ways in the process of globalisation? Conversely, to what extent do specific life course patterns actually diffuse among modern societies and, if they do, which ones are particularly successful models?

The project is organised into four phases:

- the transition from youth to adulthood (leaving the educational system, entering the job market, starting own household/family, having children, etc., and the interdependence of these events);
- mid-career men's changes in career mobility, forms of employment and unemployment over the life course;
- mid-career women's changes in career mobility, forms of employment and unemployment over the life course;
- the transition from employment to retirement.

The project uses a life course and longitudinal approach to study these processes in different countries (12-15) over longer spans of historical time, mainly the post-war period. The entire research enterprise is innovative in terms of theory (combining micro and macro approaches), methodology (using longitudinal data such panels and retrospective life histories), and in terms of a more systematic international comparison without neglecting important differences among the countries.

The Globalife project is, therefore, also expected to contribute to the application and further development of new quantitative longitudinal methodology of cross-national comparative life course research.

Source: Available from Internet: www.uni-bamberg.de/sowi/soziologie-i/globalife/index.html [cited 20.5.2004].

more likely to become unemployed than those with intermediate qualifications. A high level of qualification alone, therefore, does not improve individual labour market potential in Spain (Lemkow et al., 2000; Sokou et al., 2000).

Only a part of transition research has been carried out considering an entire life course. Studies which look at particular transition points, for example from education or initial training to work or to continuing training ignore that transitions are embedded in the framework of an individual's life history and can only be interpreted properly from this perspective.

A comparative analysis between the UK's training-on-the-job model and Germany's dual system is Hillmert (2001), using longitudinal data from the BHPS and the GLHS. Life courses and working careers of different generations (1930-70) are compared regarding transitions from education to employment, positioning in the labour market at the time of job entry and occupational career in the years after first employment. Contrary to common belief, he finds many similarities regarding the transition to work in both countries. The variation between different levels of educational attainment has increased: it becomes more and

more difficult for people with low levels of education to enter the labour market successfully. Nevertheless, in both countries intermediate vocational – non-academic – qualifications display the highest rates of direct entry into a company. Furthermore, both countries show a constant decline of the stability of the first job over the subsequent cohorts. However, the correlation of education level and successful job entry was found to be generally higher in Germany than in the UK; and does not change when looking at several subsequent cohorts. This seems to confirm the assumption that, despite attempts to improve the traditional learning-on-the-job in the UK through national VET programmes (NVQ and Modern Apprenticeship), young people still lack sufficient vocational skills. However, the British system seems to be more permeable regarding the possibility of entering a job without having the exact qualifications needed for it.

The project Globalife (Life courses in the globalisation process, see Box 4.16), funded by the German Volkswagen Foundation, analyses the influence of globalisation on different dimensions of life courses in OECD-type societies. Central to the interest of investigation is the increasing tension between the growing uncertainty of future life course situations and individuals' self-binding decisions. The analysis concentrates on the transition from youth to adulthood, changes in career mobility, forms of employment and unemployment over the life course and the transition from employment to retirement. Due to its broad research approach, it draws mainly on secondary analysis of existing longitudinal data sets rather than collecting data for itself. As there are numerous different investigations related to educational benefits in the life course under the framework of the Globalife project, it is not possible to give a summary here.

Individual life trajectories are highly determined by the educational and employ-

ment system and the links between both (Heise and Meyer, 2004). Stronger and weaker links have a major impact on the success or failure of transitions of young people into the first (stable) job. But studies of working careers indicate that the impact of education and training, and of their respective systems, goes far beyond this. Through the mediation of occupational factors, (un-) employment and social in-/exclusion, the education system contributes to the quality and quantity of the allocation of educational benefits in the whole life course (Müller et al., 2002). As the Yuseder project confirms (Starrin et al., 2000) countries with an ill-coordinated education and training system not only display low successful transitions rates but often also provide few opportunities of further education or training. Because low skilled people with low income jobs have fewer chances to access further education and training, this is likely to result in the accumulation of disadvantages throughout the life ⁽⁹⁴⁾.

4.2.1.4. *Educational benefits in a cohort perspective*

Another important topic in life course research is the investigations of cohort and generational differences of educational benefits. Life course research has rediscovered ⁽⁹⁵⁾ the relevance of the individual affiliation to a generation or age cohort as social factor in access to opportunities and gaining benefits (Mayer, 1987). Investigating the life courses of different generations/cohorts, social change over time becomes tangible; that is of special interest when considering education and training.

There are clear cohort effects ⁽⁹⁶⁾ influencing the impact of education. Subsequent birth cohorts display different education and qualification structures and are faced with different job entry opportunities resulting from economic, social and political developments ⁽⁹⁷⁾.

⁽⁹⁴⁾ In Part 3 Section 1.4, some data are presented on the participation in further education and training according to various levels of initial education.

⁽⁹⁵⁾ First attempts to look at subsequent generations (or cohorts) were mainly made by demographers and date back to the late 19th century.

⁽⁹⁶⁾ For an illustration of cohort, age and period effects in longitudinal and life course studies see Part 2 Box 2.14.

⁽⁹⁷⁾ See for example Brunello et al. (2001b); Bynner and Parsons (2001); Bender and Dietrich (2001); Bynner (1998); Hægeland et al. (1998).

Studies which compare differences in educational benefits between generations or cohorts are mostly interested in material benefits such as monetary returns, labour-market participation, career. Increasingly they also investigate non-material benefits such as family formation, health, crime and political/social participation.

Older cohorts have, on average, lower levels of qualification but had, in the economically prosperous decades after World war II, a less difficult entry to the labour market (Bynner, 1998). The economic crises in the early 1970s and 1980s changed the labour-market situation dramatically, with persistent (long-term) unemployment and increasing job competition. Simultaneously, with the educational expansion and demographic increase in most European countries in the 1960s, the first higher skilled and 'baby boom' cohorts left education and training in the 1970s. They had more difficulty in entering the labour market and obtaining skilled and stable jobs compared with previous – lower skilled and smaller sized – cohorts.

Favoured by available data sources, employment chances and benefits of different cohorts have been intensely investigated in France, Germany, the Netherlands, Norway, Sweden and the UK. Whereas Scandinavian cohort studies emphasise health related benefits and often stem from medical research (Gridley et al., 1999; Veierød et al., 1997; Hall et al., 1993), life course research in Germany, France, the Netherlands and the UK, concentrates more on material and non-material benefits⁽⁹⁸⁾. Some selected studies are summarised below

For the UK, Bynner and Parsons (2001) compare two birth cohorts (1958 and 1970) and investigate the relevance of qualifications and basic skills in protecting against unemployment. They conclude that qualifications and skills are becoming increasingly important for avoiding unemployment. Unemployment rates at different ages were consistently higher for the lower

qualified in the younger cohort than in the older cohort.

Becker (1998), on the basis of German GSOEP and GLHS data, looks at the relationship of education and life expectancy regarding successive birth cohorts and controlling for other social determinants, for example social class. Besides the fact that higher qualifications have led to an increase of the average life span in the German population, he finds that the persistence of educational inequality throughout generations has contributed to the enduring variance in life expectancy according to the social class.

In Germany, Boockmann and Steiner (2000) find declining educational returns from cohort 1925 to 1974, especially for women. Hillmert (2002), using GLHS data, detects a slightly declining effect of educational levels regarding the first job entry over four birth cohorts from 1974 to 1992. However, the advantages of vocational qualifications remain when compared with general education. Based on the same data Mayer and Brückner (1995) and Mayer (1996) compare cohorts born in 1930 and 1960. In contrast to the above findings, they find an increasing effect of education and training levels on the first job entry and an even closer link between education/training level and job status in later years (5, 10, 15) of the working career. The fact that same data may lead to different results demonstrates the importance of the method and assumptions used but also how little is known about these variables.

4.2.1.5. *Educational benefits, social background and gender*

A considerable body of research focuses on social differences concerning educational benefits, for example in terms of the influence of family background and gender on material and non-material benefits. These studies are often combined with a cohort approach (see above) and thus investigate development over time.

⁽⁹⁸⁾ For example, Konietzka (2002, 1999); Hillmert (2002); Schrijvers et al. (2001); Sackmann (2001); Born (2000); Boockmann and Steiner (2000); Becker (1998).

Concerning the influence of social background on educational benefits

Antoninis and Tsakoglou (2001: 216) find for Greece that 'the most well-off segments of the population' have the highest educational chances and benefits. As longitudinal data was not available to them they based their study on two cross-sections from the Greek household survey (HBS, 1993-94). Therefore, they could only speculate on life course effects which they estimate to be even stronger than their static results.

Lauer (2002) examines the impact of family background and gender on educational careers in France for cohorts born between 1929 and 1968. Besides a general educational upgrade throughout the cohorts, she finds the effects of parents' education on the school outcome and performance of their children to be significant. Moreover, there is also an important impact of the fathers' occupation: children of senior managers have the best educational prospects while workers offspring display the lowest educational outcomes. In most countries the disadvantages of socially deprived people cumulate rather than decline over the life course (Bynner, 1995).

Coles et al. (2002), in a review of studies on the conditions and consequences of being 'not in education, employment or training' (NEET) at the age of 16-18, find that NEET is strongly related to a poor family background. Being NEET increases the likelihood throughout the later life course of experiencing unemployment, being involved in drug or alcohol misuse or crime, having poor health and parenting at an early age, and low lifetime earnings. Persistent offending among 18-30 year olds is highly correlated with having been excluded from school, having no or low qualifications and regular drug and alcohol misuse. Evidence on persistent offenders confirms the accumulation of risk factors leading to social exclusion.

Despite numerous training and reintegration programmes for low skilled and unemployed

people, education systems still seem not to be very successful in overcoming educational inequality which is both a result and an antecedent of social inequality. However, there are differences between countries. Educational inequality related to family background seems to be stronger in those countries where traditional class structures remain, a hierarchical education and training system is prevalent and compensatory programmes are rare⁽⁹⁹⁾. But the situation is improving in many countries. Transfers in public education, the promotion of reemployment programmes and other interventions have high distributional effects and foster educational equality in these countries (Magoula and Psacharopoulos, 1999).

Concerning gender-related differences of educational benefits

Payne (1997) investigated the consequences of the choice of different post-sixteen routes in the UK, based on the YCS. She found that young women who leave full-time education at 16 are more likely than otherwise comparable young men to remain without a full-time job or training. However, young women with above average GCSE (General compulsory secondary education) results had a better chance than similarly qualified males of securing a job.

The study by Dale and Egerton (1997) looks at the educational benefits of highly educated women, based on the NCDS (the UK). Women born in 1958 are less likely to have a degree level qualification and are more likely than men to be under-performing in terms of their occupation. Both men and women receive considerable returns on their qualifications acquired. However, women earn less than men at the same level of qualification.

Born (2000), using self-collected data for Germany⁽¹⁰⁰⁾, investigates whether the choice of initial vocational training affects the life course of three subsequent female cohorts.

⁽⁹⁹⁾ See for example Schnabel and Schnabel (2002); Antoninis and Tsakoglou (2001); Tsakoglou (1993, 1997).

⁽¹⁰⁰⁾ Data are collected in the framework of research projects (1988-96) of the German special research unit Sfb (Sonderforschungsbereich) 186.

She finds that these choices significantly determine women's working careers and family life in all three cohorts. Women tend to be trained in those VET qualifications (occupations) which typically yield lower benefits. Other research studies confirm this finding. Men more frequently hold occupations with higher educational benefits in terms of income and job status than women. Moreover, as these occupations are typically attributed to male 'characteristics', women have fewer chances or are less willing to access these jobs. These occupational fields lead to higher material benefits than occupations associated with female characteristics (Born, 2000; Dale and Egerton, 1997)

Finally, women often do not have the same opportunities to participate in continuing VET programmes. The reasons are family duties or employment in those parts of the (segmented) labour market which offer fewer opportunities to access continuing VET and subsequently better career options.

4.2.1.6. *Non-material benefits*

Blaug (1992: 207) specifies that the contemporary definition of human capital theory should include both monetary and non-monetary benefits: 'people spend on themselves in diverse ways, not only for the sake of present enjoyment, but also for the sake of future pecuniary and non-pecuniary returns.'⁽¹⁰¹⁾ ⁽¹⁰²⁾. Non-material benefits are harder to measure than material ones. They include, for example, better health and life quality, citizenship and crime reduction.

A number of studies display a clear correlation (not necessarily causality) between education, training and skills on the one side and, for example, health, life quality, family formation, reduction of criminal behaviour and avoidance of social exclusion on the other

side ⁽¹⁰³⁾. Education is clearly associated with health. Becker (1998) for example finds an enduring impact of education on life expectancy when controlling for other variables. Higher qualifications have contributed – together with other measures such as improvements in preventive healthcare, medical progress, etc. – to an increase in the average life span of European populations. However, persistent educational inequalities have resulted in a considerable variance in life expectancy between social classes.

Most links between education and health, criminal behaviour and other non-material benefits are indirect (Lasheras et al., 2001; Marmot and Wilkinson, 1999; Nehru, 1993). Direct effects of education and training, such as the increase of (lifetime) earnings, the enhancement of social status and careers may also lead to the reduction of health risks, mortality and criminal behaviour. In addition, education is an important condition for the formation of cultural capital such as personal development, social participation and dealing with institutional demands (Becker, 1998). This, in turn, is essential for anticipating and processing critical life events and coping with stressors which affect life quality, self-consciousness and health.

A research project at European level concerned with non-material benefits of education and training in the life course is SEdHA (Socioeconomic determinants of healthy ageing, Box 4.17). The project is funded by the European Commission and aims to analyse socioeconomic differences in morbidity, mortality and healthy life expectancy among the elderly in different European countries. It will contribute to the explanation of these differences, by looking at risk factors and the accumulation of health

⁽¹⁰¹⁾ Non-material benefits of education and training at macrosocial level have been discussed in Chapter 2. Note that because many macrosocial benefits also apply at micro level some repetitions in this section could not be avoided.

⁽¹⁰²⁾ We should add that material benefits are more than monetary or pecuniary benefits: material benefits also include better employment opportunities or reduction of unemployment probability, occupational career advancement, etc. Of course, many of these non-monetary material benefits can be expressed or proxied by monetary terms such as income loss in case of unemployment or earnings increase of career promotion.

⁽¹⁰³⁾ For example: Blackwell and Bynner (2002); Schuller et al. (2002); Hammond (2002); Preston and Hammond (2002); Bynner et al. (2001); Lasheras et al. (2001); Hullen (1998, 2000); Becker (1998); Dale and Egerton (1997); Frenzel (1995); Blossfeld and Huinink (1990).

problems over the life-course (Mackenbach, 1999). The explanatory variable, socioeconomic status, is operationalised as educational level, and thus it is mainly the impact of education on health which will be measured. The European Community household panel (ECHP) is used to improve comparability of information along with several national surveys and longitudinal studies.

Box 4.17. Socioeconomic determinants of healthy ageing (SEdHA)

The general aim of this project is to study socioeconomic differences in morbidity and mortality among elderly people in 11 European countries and to explain these differences.

First, the study will give an overview of socioeconomic differences in morbidity in Europe by using data from the ECHP and from national surveys.

Second, mortality differences will be described by data from 10 European countries.

Third, by combining morbidity and mortality data, socioeconomic differences in healthy life expectancy will be calculated.

Fourth, socioeconomic inequalities among the elderly are explained by looking at specific diseases, specific causes of death and certain important risk factors for ill health. Finally, the effect of accumulation of disability and mortality selection over the life course on educational differences in health expectancy at older ages will be evaluated. For the last of these, data from two longitudinal studies will be used.

The final report was submitted end-2002 and has led to a number of publications.

Source: Available from Internet: www.niwi.knaw.nl/en/oi/nod/onderzoek/OND1286305/toon [cited 12.8.2004].

There are also other effects of education and training. For example, there is evidence that higher qualified people, compared with lower skilled people, are less likely to have formed partnerships. However, in contrast to general belief, for Germany at least there is evidence that although the increasing educational attainment of women has led to a delay of marriage it has not resulted in a

general decrease. But there is a strong correlation between qualification and age at birth of the first child, with the most highly qualified parents delaying parenthood longest (average age 33). There is even some indication that highly qualified women more frequently choose to have no child (Huinink, 2000; Dale and Egerton, 1997).

4.2.1.7. Biographical research

Whether or not these results from life course research can be interpreted as benefits or deficits from an individual point of view can only be explored by biographical interpretations of the individual him-/herself (Sauer-Schiffer, 2000; Rabe-Kleberg, 1994). The individual perception of educational benefits is related to educational and occupational biography and, therefore, depends on individual interpretations and self-construction rather than on objective circumstances. Biographical research uses mainly qualitative methods such as narrative studies (see Part 2, Section 2.3.2.3).

Compared with investigations from life course research there are few studies on the benefits of education and training from a biographical perspective. In addition, as the analysis of individual biographies focuses on the special, unique characteristics of each single person, it is – strictly speaking – not possible to summarise results. The aim of this kind of research is rather to come to a reasonable understanding or interpretation of life courses and their differences. The most important result from biographical research is presumably the discrepancy between objective occurrences (as investigated in life course research) and subjective interpretation of these. The same kind of experience, for example a disruption in working career, is perceived in quite different ways depending on the individual biographical background (Schaeper et al., 2000).

Many biographical studies are conducted in Ireland and in the UK⁽¹⁰⁴⁾. Keogh and Downes (1998), for example, asked 114 participants of Irish VTOS (vocational training

⁽¹⁰⁴⁾ See for example Unwin and Wellington (2001); Bloomer and Hodkinson (2000); Stone et al. (2000); Keogh and Downes (1998).

opportunities scheme) to write down their learning biography and experiences during and after having taken part in the VTOS. As well as being records of educational achievements, the stories document the move between social exclusion and inclusion, including 114 different ways to cope with the situation. While these stories generally display a very positive subjective impression of benefits from the VTOS programme, quantitative research that measures the impact of compensation-oriented vocational training schemes displays a critical picture. In this case, the results provided by quantitative research may be nuanced by the positive perceptions of participants.

The sometimes marked difference between subjective (perceived) and objective (measured) benefits of education and training might make it difficult to assess the outcomes of a programme but they will contribute to a more complete view of results, provide interpretative scenarios and might also indicate ways for improvement. For this reason Schaeper et al. (2000) complement their quantitative analysis of working careers after participation in VET by biographical interpretations of the participants. They conclude that different subjective interpretations of discontinuous careers are correlated with the profession and occupation a person has attained through VET. Among those with lower VET levels, negative interpretations (discontinuity as threat) prevail while positive interpretations (discontinuity as chance) are more often among higher level occupations.

3.1. Discussion ⁽¹⁰⁵⁾

Research studies on individual returns on education and training suggest considerable individual benefits of education, training and skills, monetary and non-monetary. The high RoR (between 6 and 17 %) on investment in education and training clearly shows that it

constitutes a more profitable investment for individuals than others such as physical capital; savings, for example, lead to return higher than 5 % only in exceptional cases. In addition, RoR have not changed very much during the past decade, despite a dramatic educational expansion. Some national data even display a widening gap between earnings of higher and lower skilled people (thus increasing rates of return). However, one must realise that this widening gap might not be the result of increased returns on education but of increased inequality in the labour market.

Furthermore, education and training at upper secondary and higher levels are most likely to increase labour force participation. This applies particularly to women as the labour force participation rate of men is already rather high. Similarly, unemployment risks are significantly negatively associated with higher skills in most countries in recent decades, again despite the dramatic increase in the supply of more highly skilled people and a complementary reduction in the lower skilled in the labour market.

In addition, education and training yield considerable non-material benefits in terms of better health, parenting, crime reduction and social inclusion. However – as was the case at the macro level (see Section 2.1) – education and training influence such benefits indirectly.

Despite these common trends, research displays considerable differences between countries in average monetary returns, employment and unemployment rates and non-material benefits ⁽¹⁰⁶⁾. The obvious suggestion for countries with lower labour-market participation and higher unemployment would be to increase and better target their investment in education and training. This is justified by research results which indicate substantial individual and social benefits from policies aimed at lowering the number of early school leavers and providing socially deprived groups with possibilities and incentives of continuing education and training (Asplund,

⁽¹⁰⁵⁾ This section draws mainly on Heise and Meyer (2004).

⁽¹⁰⁶⁾ Human capital rates of return, however, ignore grown structures of a country which may have generated inequalities in the access to, and outcomes of, education and training. Furthermore, these models mostly assume a balanced situation which can not be found in reality.

2003). These are also important goals of the European Union, formulated on its councils in Lisbon, Barcelona, Stockholm and Copenhagen within the past four years.

A relatively new area of education and training research is to look at the benefits for individuals during their life course or analysing their biographies. Compared with cross-section analyses which are based on data for one point in time, longitudinal investigations make visible contexts and dependencies of human capital formation. Human capital formation and allocation is not static but a process in the individual life course and historical change.

In Europe, longitudinal research is concentrated in Germany, France, the Netherlands, the UK, and some Nordic Member States, whereas there is still a weak database for life course investigations in southern and eastern European countries. Results from studies dealing with a broader concept of educational benefits, including different aspects of an individual's achievement in status over the life course, make it difficult to formulate general trends and recommendations, because they display very different approaches, focal points, empirical methods and measurements. Therefore, comparisons between different countries and even between different studies for a given country are difficult. Nevertheless, there are some general tendencies and conclusions for further research.

Life course and biographical studies demonstrate the relevance of education and training for the whole individual life. They show that participation in education and training and upgrading skills have increasing and cumulative effects for occupational career and personal development ⁽¹⁰⁷⁾. However, empirical life course research also shows the still existing selectivity of access to CVT which depends on previous educational level, gender and family background and other social factors. All these contribute to the persistence and accumulation of social discrimination

throughout the life course. This implies that a mere increase in continuing training opportunities would not be an appropriate solution for improving individual labour-market participation. Instead, this might also lead to undesirable side effects such as the further polarisation of educational and occupational opportunities.

Life course research also shows the 'wider' or non material benefits of education and training, especially regarding health, life expectancy, personal development and family formation. Increased awareness of these cross impacts between different social spheres – education/training, health, family, crime, etc. – would be a powerful political tool for integrating different 'competing' policy areas (and budgets).

Regarding the cumulative effects of initial and continuing VET throughout the life course, some important policy recommendations are:

- (a) to remove the obstacles to participation in education and training of socially deprived groups and to foster actively the participation of the lower qualified in continuing education programmes;
- (b) to ease transitions between education and training, initial education/training and higher education and between education/training and work;
- (c) to facilitate vertical and horizontal mobility in a system of lifelong learning.

Studies using long-term prospective panels or retrospective analyses indicate that it becomes increasingly difficult over time to ascribe specific impacts to initial education and training or a particular programme. A range of other factors such as life experience, personal attitudes, family and social background interfere with education and training.

Assessing the impact of education and training over time would be highly relevant in view of the rapid economic advancement of most European countries in recent years. This, however, requires increased efforts to set up comparable longitudinal data sets across all European countries ⁽¹⁰⁸⁾.

⁽¹⁰⁷⁾ See for example Bukodi and Roberts (2002); Noguera et al. (2002); Becker and Schömann (1999).

⁽¹⁰⁸⁾ See Box 2.16 in Part 2.

Annexes

Table of contents

1.	Key terms related to evaluation and impact of education and training	271
2.	Comparing microeconomic methods of evaluation: a numeric example	274
3.	Changing contexts and drivers for change for education and training: detailed tables and figures	277
4.	Selected results and recommendations from the OECD thematic review on adult learning	286
5.	Evaluation associations	289
6.	Journals on evaluation in Europe and beyond: alphabetical list	291

List of tables and figures

Tables

A1.	Labour-market outcomes for participants and non-participants (example)	274
A2.	Estimates of the programme effects	276
A3.	Working age population, 25-64 years old, baseline scenario until 2040 and average in 2001, EU15	277
A4.	Ageing of the EU working age population: ration of persons aged 45-64/25-44, EU15 and accession countries	278
A5.	Comparison of the education level of two generations, EU15 and accession countries, 2002 (% of respective age group)	279
A6.	Population of 18 to 24 years old having left the education system without qualifications (ISCED 0-2), 2002, 2000, compared with earliest data available (% of the respective age group)	280
A7.	Participation rates in lifelong learning, 25-64 years old, by purpose of training, 2002 (%)	281
A8.	Willingness to pay for education and training, by purpose, population aged 15 and more, EU15, 2003 (%)	282
A9.	Employment rates, 25-64 years old, by education level, EU15 and accession countries, 2002 (%)	283
A10.	Unemployment rate, 25-64 years old, by educational level, 2002 (%, sorted by total)	285
A11.	Recommendations to enhance access and participation in adult learning: five key areas	288

Figures

A1.	Average population 2001, distribution by age group, EU15 and accession countries (sorted by size of the younger age group)	278
A2.	Comparison of the education level of two generations, by gender, EU15, 2002 (%)	280
A3.	People's opinion on the usefulness of the 'learning to learn' competence, comparison of generations, EU15 plus Iceland and Norway, 2003	282
A4.	People's opinion on the usefulness of the 'learning to learn' competence, by country, 2003 (sorted by usefulness in private life)	283
A5.	Inactive people, 25-64 years old, by gender and education level, EU15, 2002 (%)	284
A6.	Inactive people, 25-64 years old, by gender and education level, accession countries, 2002 (%)	284

ANNEX 1

Key terms related to evaluation and impact of education and training ⁽¹⁾

Assessment

DE: *Bewertung*; **FR:** *évaluation*

A general term embracing all methods used to appraise/judge performance of an individual, a group or a programme.

Comment:

- (1) in English, the term 'evaluation' generally refers to the assessment of training programmes and policies;
- (2) 'assessment of competences' (*Bewertung von Kompetenzen / Evaluation des compétences*) refers to the sum of methods and processes used to evaluate the attainments (knowledge, know-how and/or competences) of an individual, and typically leading to certification.

Source: Tissot, 2004.

Related term: evaluation

Benchmarking

DE: *Richtwert, Bezugswert, Maßstab, Benchmarking*;

FR: *étalonnage des performances, évaluation critique*

A systematic process comparing the activities, processes and/or performance of a programme, organisation, country, etc. against a theoretical, political or existing reference to identify ways to improve performance.

Comment: benchmarking can be either qualitative or quantitative.

Source: adapted from The Economist, Business Database Dictionary. Available from Internet: <http://www.economist.com/businessDatabase/Dictionary.cfm?id=ID8D747D13-8E5B-11D5-8499-00508B2CCA66> [cited 24.5.2004].

Benefit (see 'outcome')

DE: *Vorteil, Nutzen*;

FR: *Utilité avantage, bénéfice, gain*

Effect

DE: *Wirkung, Effekt*; **FR:** *effet*

Socioeconomic change resulting directly or indirectly from an intervention.

Comment: Effects include the results and impacts of an intervention, whether positive or negative, intended or not.

Source: adapted from European Commission 1999, Means collection, Vol. 6.

Effectiveness

DE: *Effektivität, Wirksamkeit*; **FR:** *efficacité*

The extent to which the objectives of a policy or an intervention are achieved, usually without reference to costs.

Source: Tissot, 2004;
see also: ISO 9000-2000.

(¹) This glossary was prepared by Philippe Tissot, terminologist at Cedefop, assisted by Alena Zuckersteinová.

Efficiency

DE: *Effizienz, Kosten-Wirksamkeit;*

FR: *efficience*

Relationship between the results achieved (output) and the resources used (input).

Source: ISO 9000-2000.

Evaluation

DE: *Evaluation, Evaluierung;* **FR:** *évaluation*

Judgment on the value of an intervention, programme or policy with reference to criteria and explicit standards (e.g. its relevance, its efficiency).

Comment: the generic definition above does not explicitly cover all aspects reviewed in detail in the present report. Evaluation encompasses two broad aspects:

- (1) evaluation as a systematic investigation to determine the worth or merit of a programme, measure or policy by means of careful appraisal and study, based on relevant social research methods and criteria, standards and indicators (summative or impact evaluation);
- (2) evaluation as a developmental process that illuminates or enlightens specific policies, processes and practice for its stakeholders, contributes to collective learning, reduces uncertainty in decision making and helps to improve the design and implementation of the programme and/or of future related initiatives (formative or process evaluation).

Sources:

- (1) European Commission 1999, *Means collection*, Vol. 6 (definition);
- (2) Tissot, 2004 (comment).

Related terms: assessment, *ex ante* evaluation, *ex post* evaluation, formative evaluation, summative evaluation.

Ex ante evaluation

DE: *Ex-ante Evaluierung;*

FR: *évaluation ex ante*

An evaluation conducted before the implementation of an intervention to provide a prior assessment of whether socioeconomic issues have been diagnosed correctly, whether the strategy and objectives are relevant, whether it is coherent with other interventions or policies, etc.

Source: adapted from

- (1) UK Evaluation Society, Glossary of evaluation terms. Available from Internet: http://www.evaluation.org.uk/Pub_library/Glossary.htm [cited 20.5.2004];
- (2) European Commission 1999, *Means collection*, Vol. 6.

Syn.: preformative evaluation.

Related terms: assessment, evaluation, *ex post* evaluation, formative evaluation, summative evaluation.

Ex post evaluation

DE: *Ex-post Evaluierung;*

FR: *évaluation ex post*

An evaluation conducted either on or after completion of an intervention or programme. It aims at accounting for the use of resources, effectiveness and efficiency of intervention and strives to understand the factors of success or failure.

Source: adapted from

- (1) UK Evaluation Society, Glossary of evaluation terms. Available from Internet: http://www.evaluation.org.uk/Pub_library/Glossary.htm [cited 20.5.2004];
- (2) European Commission 1999, *Means collection*, Vol. 6.

Related terms: assessment, evaluation, *ex ante* evaluation, formative evaluation, summative evaluation.

Formative evaluation

DE: *Gestaltungsevaluation, formative Evaluation*;

FR: *évaluation formative*

An evaluation examining ways of improving and enhancing the implementation and management of interventions. Formative evaluations are primarily conducted for the benefit of those managing the intervention with the intention of improving their work or related future initiatives.

Source: adapted from UK Evaluation Society, Glossary of evaluation terms. Available from Internet: http://www.evaluation.org.uk/Pub_library/Glossary.htm [cited 20.5.2004];

Syn.: process evaluation

Related terms: assessment, evaluation, *ex ante* evaluation, formative evaluation, summative evaluation.

Impact

DE: *Wirkung*; **FR:** *impact*

A general term used to describe the effects of a programme, policy or socioeconomic change. Impacts can be positive or negative, direct or indirect, as well as foreseen or unforeseen.

Comment: impact research analyses the effects resulting from an intervention or programme or from changes in society such as demographic or technological change, educational expansion, etc.

Source: adapted from UK Evaluation Society, Glossary of evaluation terms. Available from Internet: http://www.evaluation.org.uk/Pub_library/Glossary.htm [cited 20.5.2004];

Related terms: output, outcome.

Input

DE: Input; **FR:** *intrans*

The human, financial and physical resources used for an intervention.

Source: adapted from UK Evaluation Society, Glossary of evaluation terms. Available from Internet: http://www.evaluation.org.uk/Pub_library/Glossary.htm [cited 20.5.2004].

Outcome

DE: Outcome, *Ergebnisse*;
FR: *résultat, bénéfice*

The positive or negative longer term socioeconomic changes or impacts that occur directly or indirectly as a result of an intervention's inputs, activities and outputs.

Source: adapted from *Accountability and evaluation glossary*. Available from Internet: <http://www.nonprofitbasics.org/TopicArea/Glossary.aspx?ID=14> [cited 20.5.2004].

Output

DE: Output; **FR:** *extrant*

Immediate and direct tangible result of an intervention.

Source: Tissot, 2004.

Related term: outcome.

Summative evaluation

DE: *Bilanzevaluation, summative Evaluation*;
FR: *évaluation sommative*

A systematic investigation to determine the worth or merit of a programme, measure or policy using relevant social research methods and criteria, standards and indicators.

Source: Part 1, Section 1.2 of this report.

ANNEX 2

Comparing microeconomic methods of evaluation: a numeric example

Hujer et al. (2004a, Section 2.2.7) present a hypothetical example to illustrate the differences in evaluation results depending on the microeconomic method used (see Part 2, Chapter 2 of this report). In their example, under evaluation is a programme that intends to improve management skills and employment prospects. The evaluators decided that the outcome category is ‘employed’ (Y , yes = 1; no = 0). To underline the role of heterogeneity (see Section 1.3.2.2, Part 2) it is assumed that there are only two types of worker – high-skilled and low-skilled – and that the programme works better for high-skilled workers.

Table A1. Labour-market outcomes for participants and non-participants (example)

Non-participants			Participants			
Low-skilled workers						
Worker	Y_{t-1}	Y_{t+1}	Worker	Y_{t-1}	Y_{t+1} (unobserved counterfactual outcome)	$Y_{t+1} + T$ (actual outcome)
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1	0	1	11	0	0	1
2	0	0	12	0	0	0
3	0	0	13	0	0	0
4	0	0	14	0	1	1
5	0	0	15	0	1	1
6	1	1	16	0	0	0
7	1	1	17	0	1	1
8	1	1	18	1	1	1
9	1	1	19	1	1	1
10	1	1	20	1	0	1
Average	0.5	0.6		0.3	0.5	0.7
High-skilled workers						
21	0	0	31	0	0	1
22	0	1	32	0	1	1
23	0	0	33	0	0	1
24	1	1	34	1	1	1
25	1	1	35	1	1	1
26	1	1				
27	1	1				
28	1	1				
29	1	1				
30	1	1				
Average	0.7	0.8		0.4	0.6	1

NB: Y refers to the outcome (employed: yes or no); $t-1$ refers to the period before the programme started, $t+1$ to the period after the programme; T is the observed treatment effect.

Source: Adapted from Hujer et al. 2004a, Section 2.2.7.

Table A1 displays the labour-market outcomes of those workers who participated in the programme and those who did not in the period t . Columns 2 and 3 show the employment situation Y of non-participants before (Y_{t-1}) and after (Y_{t+1}) the programme. Column 5 contains the pre-programme employment situation of participants (Y_{t-1}) and column 7 contains their post-programme situation ($Y_{t+1} + T$), where T is the ‘true’ treatment effect. All these outcomes are actually observed. Column 6 contains the counterfactual (hypothetical) employment situation without treatment for participants. This unobservable outcome is only introduced as a reference level to illustrate the functioning of the different estimators. Clearly, this outcome is never observed in an empirical study but allows us to answer the question ‘What would have happened to the participants if they had not participated?’ The ‘true’ programme impact can be calculated by comparing the actual situation of participants after the programme took place ($Y_{t+1} + T$) with their hypothetical counterfactual situation without treatment (Y_{t+1}). Again it is important to note that a researcher will not be able to do this since column six is unobservable. This unobservable outcome is only introduced as a reference level to illustrate the functioning of the different estimators.

The two skill groups will be discussed separately.

The low skilled workers

The group of low-skilled workers is divided into 10 non-participants (workers 1-10) and 10 participants (workers 11-20). Half of the non-participants (workers 6-10) are employed in both periods and for worker 1 the labour-market situation improves over time. In the group of participants, on the other hand, only three workers are employed in the first period (workers 18-20). For three other workers (workers 14, 15 and 17) the situation would have improved even in absence of the training

programme (column 6), whereas for one worker (No 20) it would have worsened. The ‘true’ programme impact can be estimated by comparing columns 7 and 6. The programme improves the employment prospects of two workers (No 11 and 20); therefore the ‘true’ effect is 0.2. Below is shown how a researcher would calculate the programme impacts with some of the methods presented in Section 1.3.3 of Part 2 ⁽²⁾.

The before-after method (BAE) relies on the assumption that the pre-training outcome of the participants represents a valid description of their unobservable counterfactual outcome in the post-training period $t+1$. Therefore, the estimator is calculated by comparing the outcome in column 7 with the outcome in column 5 and leads to a result of 0.4; i.e. 4 out of 10 participants have improved their employment situation. Clearly, this estimator attributes all improvements in the employment situation between the two periods to the programme. But since, in our example, the situation of two workers would have improved anyway, for example because of an overall economic improvement, the before-after method overestimates the true impact.

The difference-in-differences method (DiD) eliminates common time trends by subtracting the before-after change in non-participant outcomes from the before-after change for participant outcomes. In our example we form simple averages over the group of participants and non-participants. We then contrast observed changes in the labour-market situation of participants (columns 7–5) with the corresponding changes for non-participants (columns 3–2). The estimated impact in the example is then: $(0.7 - 0.3) - (0.6 - 0.5) = 0.3$. The DiD method overestimates the true programme effect in this example.

The cross-section method compares the average of the observed outcome of the non-participants with the average of the unobservable outcome of participants. Therefore, this estimator is calculated by

⁽²⁾ For the sake of simplicity this numerical example has been kept on a basic methodological level. Therefore, more complex estimators could not be illustrated.

comparing the actual employment situation ($Y + T$) of participants (workers 11-20) in the period $t+1$ (average: 0.7, column 7) with the actual average situation of non-participants (workers 1-10) in the same period (average: 0.6, column 3). The estimated impact thus would be $0.7 - 0.6 = 0.1$. In this example, the cross-section method underestimates the true impact.

An exact matching approach replaces the counterfactual outcome of the participants with the population average of a matched control group. In this example there is only one available matching variable, namely the labour-market situation before training, in $t-1$. Estimation is then simply done by comparing the employment situation in $t+1$ of workers 11-17 (column 7) with the situation of workers 1-5 (column 3) on the one hand and between workers 18-20 (column 7) and 6-10 (column 3) on the other. The estimated impact is then a weighted average between both groups. The effect in the first group, with seven participants who have been unemployed before training, is: $(4/7 - 1/5) = 0.57 - 0.2 = 0.37$. The effect in the second group where three workers have been employed before training is $(5/5 - 3/3) = 0$. The weighted average is therefore $(0.37 \cdot 7 + 0 \cdot 3) / 10 = 0.26$; the true effect is slightly overestimated.

Highly skilled workers

Same calculations can be done for the high-skilled workers. This group contains 10 non-participants (workers 21-30) and five participants (workers 31-35). 60 % of the participants are unemployed before training and only one participant (worker 32) would have experienced an improvement in his labour-market situation without training. The 'true' programme effect in the group of high-skilled workers is 0.4. The evaluation estimators can be calculated analogously as described above and the results can be found in column 3 of Table A2. The matching estimator is very successful this time as it exactly calculates the true impact, whereas the

'before-after' and DiD overestimate the programme impact and the cross-section method underestimates it.

Table A2. **Estimates of the programme effects**

Method	Low-skilled	Highly-skilled	Average
'True' impact	0.20	0.40	0.27
Before-after	0.40	0.60	0.47
Difference-in-differences	0.30	0.50	0.37
Cross-section	0.10	0.20	0.13
Matching	0.26	0.40	0.31

Source: Adapted from Hujer et al., 2004a, Section 2.2.7.

The last column is a weighted ⁽³⁾ average of the programme effects for the two groups. This is to illustrate how important it is to account for heterogeneity in estimating programme impacts. Due to data limitations, many evaluation studies have to pool heterogeneous programmes and/or heterogeneous individuals and estimate one composite treatment effect. It becomes clear in the given example that this approach is misleading. The weighted 'true' impact for low- and high-skilled workers taken together is 0.27. However, the true impact for high-skilled (0.4) and for low-skilled workers (0.2) differ substantially.

Disregarding the heterogeneity of participants thus leads to underestimation (overestimation) of the true effect. Clearly this is an unnecessary source of bias which has to be avoided. Therefore, evaluations should be carried out separately with respect to major individuals' characteristics, regional aspects and also programme differences.

(³) Weighted with the respective number of participants (10 low-skilled and 5 high-skilled workers).

ANNEX 3

Changing contexts and drivers for change for education and training: detailed tables and figures

Table A3. Working age population, 25-64 years old, baseline scenario until 2040 and average in 2001, EU15

	2000 (x1000)	Recorded population average 2001 (x1000)	2000 = 100			
			2010	2020	2030	2040
EU15	204 914	:	102.2	100.9	95.2	88.3
B	5 476	5 499	102.8	101.2	94.6	89.5
DK	2 939	2 964	100.2	97.9	97.1	92.4
D	46 786	46 446	97.2	97.0	89.0	81.9
EL	5 644	:	105.4	103.8	99.5	92.3
E	21 118	21 970	107.6	104.9	97.5	86.6
F	30 914	30 815	105.7	104.0	100.5	96.8
IRL	1 868	1 939	119.7	126.0	128.8	129.9
I	32 132	:	99.3	94.5	86.3	74.8
L	240	246	105.7	109.9	110.3	109.1
NL	8 882	8 969	103.5	102.2	99.5	95.2
A	4 523	4 562	100.6	101.3	93.7	84.7
P	5 306	5 495	107.2	106.8	106.0	101.8
FIN	2 802	2 815	102.7	97.0	91.8	89.1
S	4 665	4 709	101.0	102.2	99.4	95.2
UK	31 618	:	103.2	104.3	100.6	95.2

NB: 2000 = 100%; ':' not available.

Source: Eurostat, projections 1995, revised 1999 and population average, Newcronos.

Table A4. Ageing of the EU working age population: ratio of persons aged 45-64/25-44, EU15 and accession countries

	Ratio (2001) recorded	Ratio (2040) baseline scenario
EU15	0.79 ^(a)	1.09
B	0.82	1.08
DK	0.88	0.96
D	0.85	1.11
EL	0.82 ^(a)	1.14
E	0.69	1.20
F	0.83	1.02
IRL	0.71	1.05
I	0.81 ^(b)	1.16
L	0.72	0.93
NL	0.80	0.96
A	0.76	1.13
P	0.80	1.03
FIN	1.00	1.10
S	0.93	1.13
UK	0.78 ^(b)	1.05
CY	0.76 ^(b)	
CZ	0.93	
EE	0.89 ^(b)	
HU	0.96	
LT	0.76	
LV	0.87	
MT	0.92	
PL	0.84	
SI	0.84	
SK	0.77	

(a) 1999; (b) 2000.

Source: Newcronos, Eurostat, Demography, population average and baseline scenario 1999, calculations by authors.

Figure A1. Average population 2001, distribution by age group, EU15 and accession countries (sorted by size of the younger age group)

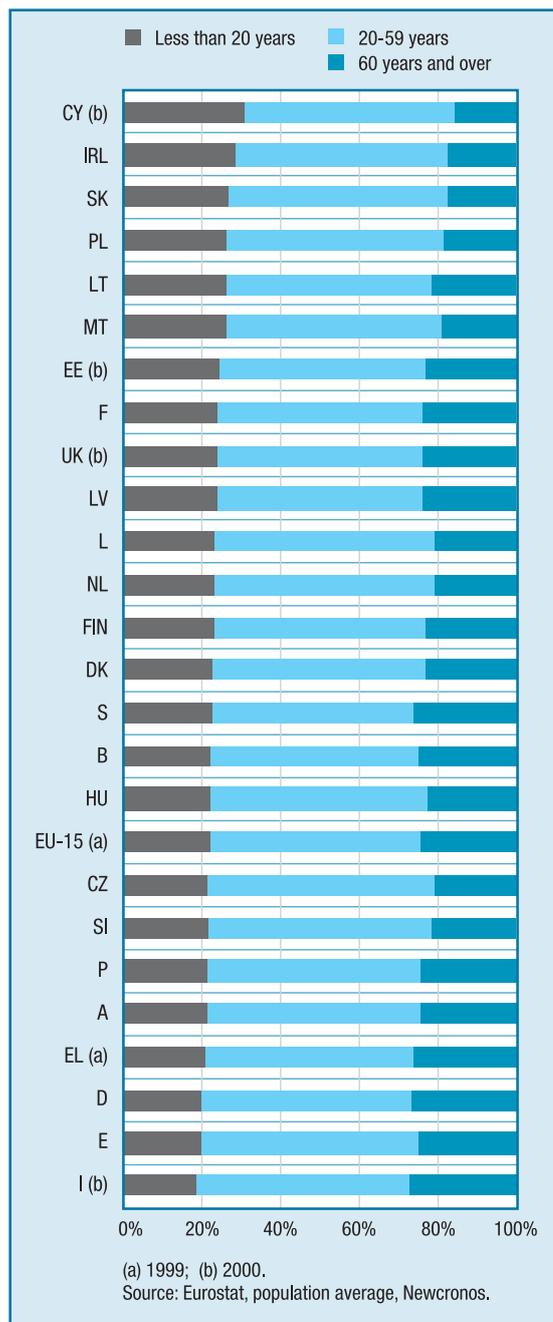


Table A5. **Comparison of the education level of two generations, EU15 and accession countries, 2002 (% of respective age group)**

	55-59			30-34		
	Isced 0-2	Isced 3-4	Isced 5-6	Isced 0-2	Isced 3-4	Isced 5-6
EU15	47.8	35.1	17.1	26.7	47.3	26.0
B	54.3	25.3	20.3	25.9	39.1	34.9
DK	22.5	50.0	27.5	14.3	54.3	31.4
D	20.1	58.6	21.3	14.4	61.4	24.2
EL	68.5	21.1	10.4	28.3	47.7	24.1
E	79.8	7.9	12.3	45.3	21.7	33.0
F	49.9	33.0	17.1	23.4	45.0	31.5
IRL	69.9	12.0	18.1	33.9	24.8	41.3
I	72.8	19.3	7.9	45.0	41.8	13.3
L	47.8	39.1	13.0	29.7	45.9	24.3
NL	44.5	34.5	21.0	23.7	48.0	28.3
A	29.1	56.5	14.3	16.1	65.9	17.9
P	90.5	4.4	5.1	71.2	15.9	12.8
FIN	42.0	31.5	26.5	13.9	44.9	41.3
S	29.3	46.1	24.6	9.0	62.7	28.3
UK	32.2	44.9	22.9	11.2	58.4	30.3
CY	57.1	28.6	14.3	16.3	46.9	36.7
CZ	16.8	71.0	12.2	5.6	81.9	12.5
EE	23.5	50.0	26.5	– ^(a)	62.6	28.6
HU	40.7	45.6	13.8	20.1	66.0	13.9
LT	30.6	31.7	37.8	4.7	49.6	45.7
LV	33.3	50.0	16.7	13.0	69.6	17.4
PL	36.4	53.7	9.9	10.6	75.6	13.8
SI	31.2	56.9	11.9	17.7	62.4	19.9
SK	25.1	65.5	9.4	8.2	81.7	10.1

NB: No data available for Malta;

(a) not published because of too small sample size.

Source: Newcronos, Eurostat, Labour force survey, calculations by authors.

Table A6. Population of 18 to 24 year olds having left education without qualifications (ISCED 0-2), 2002, 2000, compared with earliest data available (% of the respective age group)

	1992	2000	2002
B	18.1	12.5	16.3
DK	15.2	11.6	19.6
EL	25.2	17.1	19.5
E	40.4	28.8	30.2
F	17.2 ^(a)	13.3	13.4
IRL	27.1	–	27.1
I	37.7	25.3	23.8
L	42.2	16.8	18.1 ^(b)
P	50.0	43.1	45.5
UK	34.7	6.8	6.9

	1996	2000	2002
D	13.3	14.9	12.5 ^(b)
NL	17.6	15.5	18.0
A	13.6 ^(c)	11.9	9.5
FIN	11.1	8.9	10.7
S	7.5	7.7	16.2
SI	10.8	7.4	4.9

	1999	2000	2002
CY	15.1	15.0	14.0
CZ	–	–	6.0
EE	14.0	14.2	13.6
HU	13.0	13.8	12.4
LT	27.0	17.2	13.2
LV	–	–	17.6
PL	–	6.6 ^(b)	7.6
SK	–	–	5.3

NB: No data available for Malta;
(a) 1993; (b) 2001; (c) 1995.

Source: Newcronos, Eurostat, Labour force survey.

Figure A2. Comparison of the education level of two generations, by gender, EU15, 2002 (%)

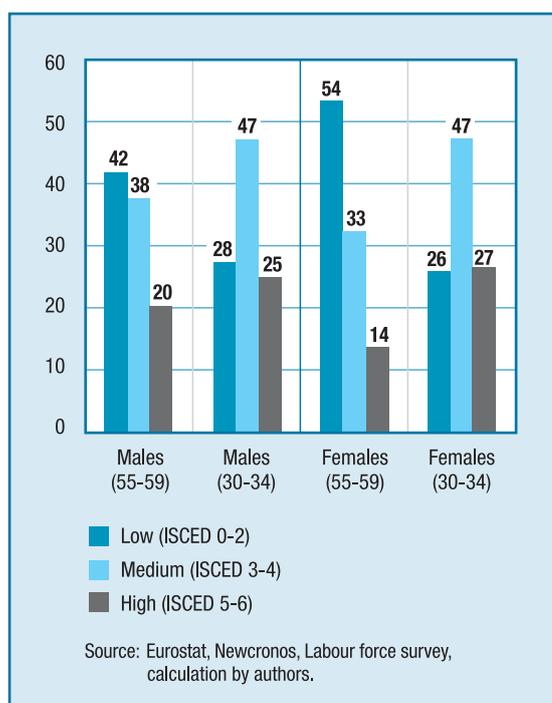


Table A7. **Participation rates in lifelong learning ^(a), 25-64 years old, by purpose of training, 2002 (%)**

	Lifelong learning total	Purpose of training			
		To adapt to technological change, obtain promotion or upgrade skills	Training under specific employment measure	General interest	Not known
EL	0.4	0.3	:	:	:
F	2.0	2.0	:	:	:
P	2.3	0.7	–	0.9	0.3
E	3.1	0.8	0.0	2.2	:
D ^(b)	3.9	2.6	0.6	0.5	0.2
L ^(b)	4.1	2.9	:	0.8	:
I	4.2	0.2	1.1	0.1	2.8
B	5.4	0.2	2.3	2.9	:
A	5.8	3.3	0.5	2.0	:
IRL	6.1	4.5	0.4	1.3	:
DK	12.3	8.5	1.7	1.7	0.4
S	14.6	7.4	0.9	3.3	2.9
NL	15.4	8.4	0.4	6.6	:
FIN	15.8	13.1	0.7	1.9	:
UK	20.4	11.1	0.3	:	9.0
EE	2.4	2.1	:	:	:
LT	2.4	1.4	:	0.8	:
HU	2.7	2.3	0.1	0.3	:
PL	3.0	1.6	–	0.2	1.2
CY	3.0	2.5	:	0.8	:
SI	4.1	4.1	:	:	:
CZ	4.5	1.1	0.8	2.5	0.0
LV	5.1	3.9	:	1.2	:
SK	8.4	3.0	0.2	5.2	:

NB: No data available for Malta;

(a) excluding initial ET; (b) 2001;

– = not published because of too small sample size;

: = not available.

Source: Newcronos, Eurostat, Labour force survey, calculations by authors.

Table A8. Willingness to pay for education and training, by purpose, population aged 15 and more, EU15, 2003 (%)

To...	keep present job	have a better private life	get a promotion	learn a new language	set up one's own business	get new knowledge for a hobby	open up job and career opportunities	get a recognised certificate	get a pay rise	prepare for retirement	get new knowledge in one's work field	get back into the labour market
would contribute financially	37.7	51.4	38.7	46.7	45.2	46.3	48.2	48.1	39.7	34.8	43.6	39.7
would pay all of the cost	12.9	21.8	11.7	18.6	23.0	21.5	16.4	18.1	14.8	11.7	12.9	14.8
would pay some of the cost	24.8	29.6	27.0	28.0	22.2	24.8	31.8	30.0	24.9	23.1	30.6	24.9
would pay none of the cost	46.7	39.1	48.7	44.9	41.2	44.4	40.5	42.2	47.4	50.9	46.0	45.5
do not know	2.7	9.5	12.6	8.4	13.6	9.3	11.3	9.7	12.9	14.3	10.4	14.8

Source: Lifelong learning questionnaire included in standard Eurobarometer, wave 59.0, 2003.

Figure A3. People's opinion on the usefulness of the 'learning to learn' competence, comparison of generations, EU15 plus Iceland and Norway, 2003

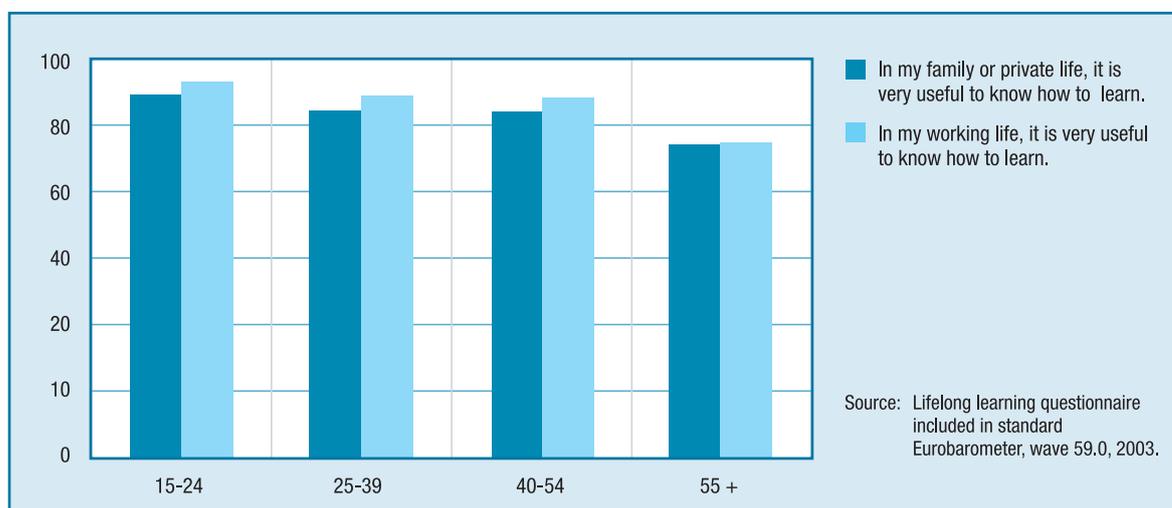


Figure A4. People's opinion on the usefulness of the 'learning to learn' competence, by country, 2003 (sorted by usefulness in private life)

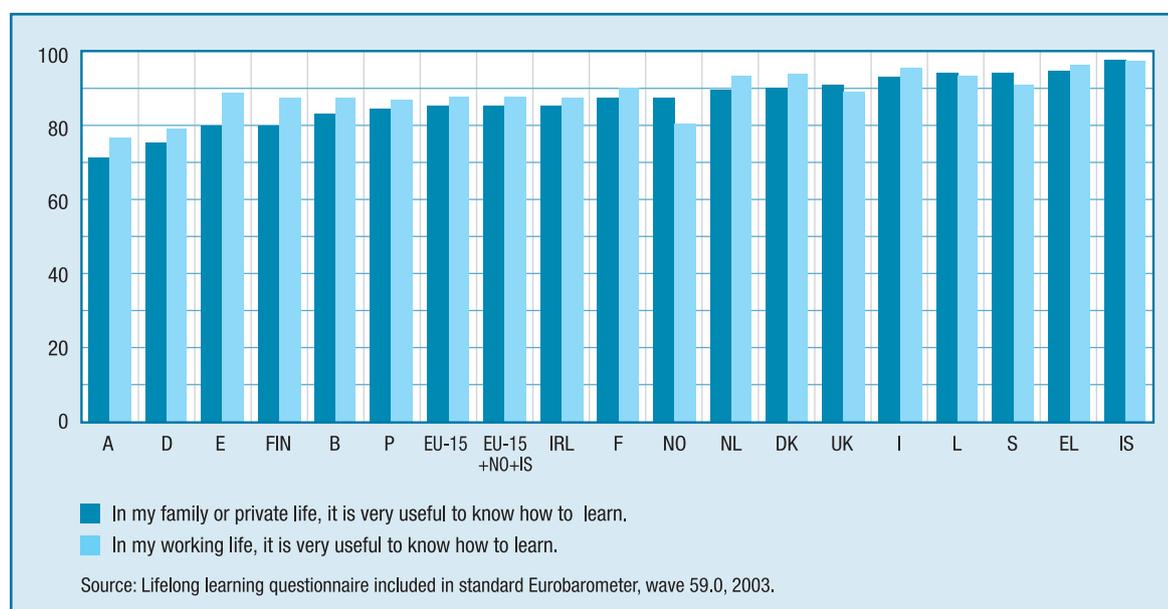


Table A9. Employment rates ^(a), 25-64 years old, by education level, EU15 and accession countries, 2002 (%)

	ISCED 0-2	ISCED 3-4	ISCED 5-6		ISCED 0-2	ISCED 3-4	ISCED 5-6
EU15 ^(b)	54.9	74.6	84.5	FIN	58.7	75.5	85.6
B	48.2	74.7	83.6	S	68.3	82.1	87.6
DK	61.0	81.5	87.2	UK	54.2	80.7	87.9
D ^(b)	52.3	70.3	83.3	CY	64.0	77.8	88.0
EL	55.9	65.7	81.2	CZ	45.9	76.2	87.4
E	55.7	71.7	80.9	EE	48.1	72.6	80.2
F	56.8	76.1	83.5	HU	36.8	71.6	82.1
IRL ^(b)	56.8	77.2	87.1	LT	42.0	71.6	78.9
I	50.5	72.4	82.3	LV	48.8	71.2	81.7
L ^(b)	58.2	74.2	85.5	PL	38.5	62.7	83.5
NL	60.7	79.8	86.7	SI	57.6	74.7	86.7
A	53.8	73.9	85.6	SK	28.6	70.0	86.9
P	73.3	82.0	89.6				

NB: No data available for Malta;

(a) percent of the respective age group in employment; (b) 2001 data.

Source: Newcronos, Eurostat, Labour force survey.

Figure A5. Inactive people, 25-64 years old, by gender and education level, EU15, 2002 (%)

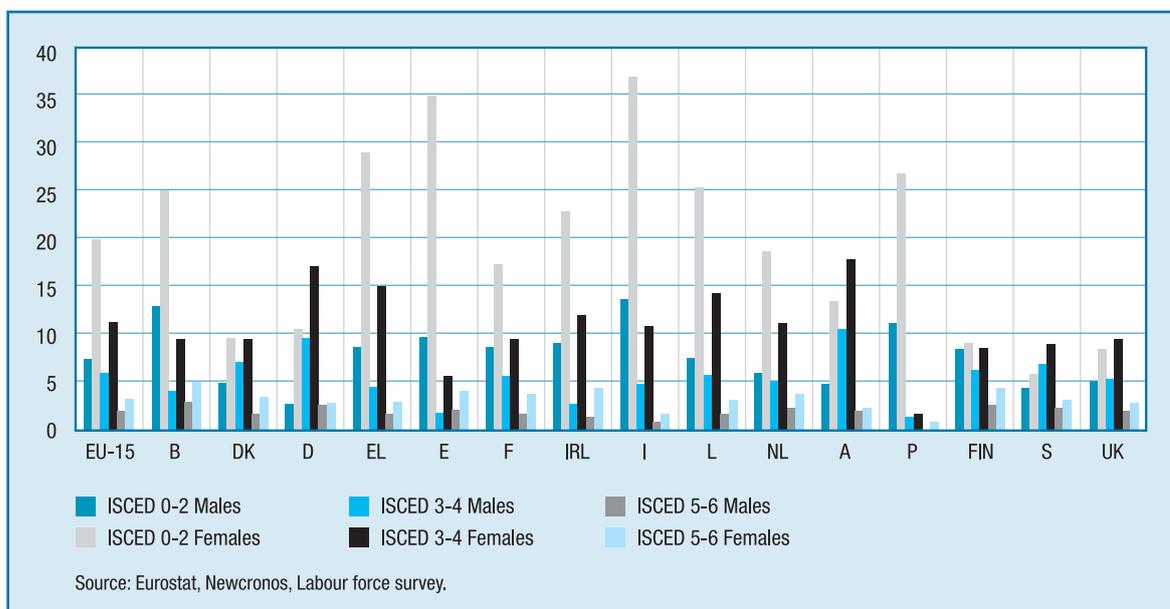


Figure A6. Inactive people, 25-64 years old, by gender and education level, accession countries, 2002 (%)

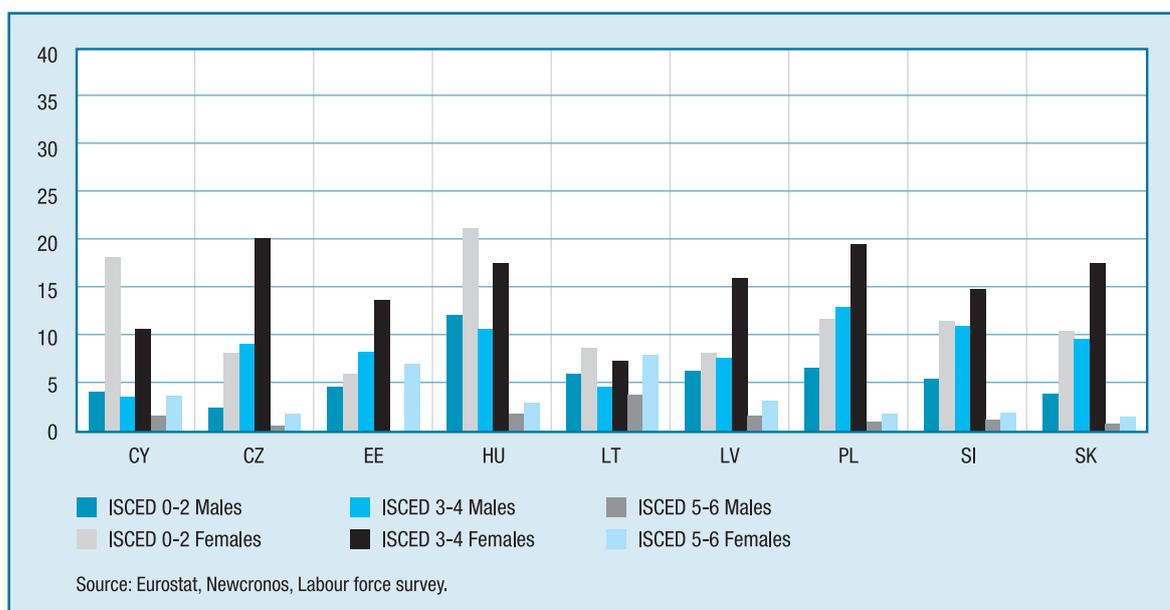


Table A10. **Unemployment rate ^(a), 25-64 years old, by educational level, 2002
(%, sorted by total)**

	Total	Low (ISCED 0-2)	Medium (ISCED 3-4)	High (ISCED 5-6)		Total	Low (ISCED 0-2)	Medium (ISCED 3-4)	High (ISCED 5-6)
L	2.2	3.8	(1.2)	(1.8)	EU15	6.7	9.5	6.4	4.2
NL	2.2	2.9	1.9	1.7	I	7.4	9.0	6.4	5.3
CY	2.9	3.4	3.2	1.9	FIN	7.4	11.0	8.5	4.0
IRL	3.6	5.9	2.9	1.9	F	7.5	11.3	6.5	5.0
DK	3.9	6.2	3.4	3.5	EL	7.9	7.3	9.6	6.3
P	3.9	4.0	4.2	3.3	D	8.5	15.0	8.7	4.3
S	4.0	5.3	4.3	2.6	EE	8.8	–	10.0	–
UK	4.0	8.1	3.7	2.4	E	9.7	11.1	9.4	7.4
A	4.6	8.2	4.5	1.7	LV	11.8	19.7	12.0	6.2
HU	5.0	10.5	4.4	1.5	LT	12.4	17.0	15.7	8.5
SL	5.0	7.8	5.0	2.2 ^(b)	SK	15.4	42.5	14.3	3.1
B	6.0	9.8	5.6	3.1	PL	16.9	25.5	17.6	5.6
CZ	6.0	18.0	5.5	1.6					

NB: No data available for Malta;

(a) % of labour force (employed + unemployed) of the respective age group; (b) unreliable or uncertain data;
– not published because of too small sample size.

All data correspond to the second quarter of 2002, except France which corresponds to the first quarter.

Source: Newcronos, Eurostat, Labour force survey.

ANNEX 4

Selected results and recommendations from the OECD thematic review on adult learning ⁽⁴⁾

All the countries which participated in the OECD thematic review (Canada, Denmark, Finland, Norway, Portugal, Spain, Sweden, Switzerland and the UK) have taken policy measures targeted towards adult learning at a national level. Most governments believe that such intervention is justified.

Some countries have a somewhat fragmented vision of adult education policies, separating educational and vocational policies under the auspices of different ministries (as in Spain) or attributing them different objectives depending on the stakeholders (as in Portugal). Others, placing adult learning in the broader framework of lifelong learning, have more holistic policies targeted at all adults (Denmark, Finland, Norway, Sweden and the UK).

Although the objective of upskilling and reskilling adult population is common to most country policies (including the support to basic skills instruction, either of primary or secondary level), different approaches are being taken (separately or in combination):

- (a) stimulating demand through financial incentives targeted at individuals (Canada, Finland, Norway, Sweden and the UK) or through non-financial incentives in the form of education and training leave for example (Finland, Norway and Portugal);
- (b) recognising non-formal and informal learning to bring adults back to learning and shorten the learning process. Most countries have been initiating plans or activities in this respect. The Swedish government has begun

developing a general system for assessment of competences, assisted by the social partners. In Canada, prior learning assessment is institutionalised across the country. In Denmark, the Adult Education Reform foresees the creation of a coherent recognition system under the ministry of labour. In Norway, one of the key elements of the Competence Reform is a national system for the documenting and recognising non-formal learning. Also in Portugal, the national system for recognising, validating, certifying school attainment and personal experience includes developing key competence benchmarks ⁽⁵⁾;

- (c) stimulating the adult learning market through guidance and counselling (such as in Finland and Switzerland), quality assurance schemes (Switzerland and the UK), subsidies to private suppliers (Denmark), etc.;
- (d) developing learning in enterprises through a tax levy for funding (Canada [Quebec], Spain, Switzerland [Canton of Geneva]) or for compensating income loss for employees (Finland). There are also tax exemption (Norway); publicly financed training leave (Denmark); entitlement to minimum amount of training and minimal annual volume of training (Portugal).
- (e) increasing and rationalising supply (reforms in Denmark, Finland, Norway, Sweden and efforts undertaken in Spain and Portugal). In this respect, national qualification frameworks, offering the

⁽⁴⁾ Annex 2 is based on OECD, 2002 and 2003a: 69-101.

⁽⁵⁾ For an in-depth discussion of the reform of certification systems, assessment and recognition of skills, the reader may refer to Descy and Tessaring (2001a: 77-87).

possibility of various progression routes, equivalences, credit transfer and other kinds of bridges between pathways, are one way of developing more open and effective systems. Examples of such frameworks are the Finnish qualifications framework relying on competence-based examinations, the NVQ in the UK or modularised systems such as the ones introduced in Denmark and Switzerland. Spain and Portugal are also in the process of defining national qualification frameworks.

A general trend in policies encountered is the attempt to be more responsive to demand rather than being strictly supply-led. The decentralisation of decision-making is a characteristic mechanism in this respect, as is the wider use of direct incentives to individuals.

Countries have designed different solutions to compensate for the inequities in access to adult learning: targeted programmes of spending, subsidies for groups with special needs, special outreach programmes, and various initiatives designed to remove barriers to participation (such as child care). Although the specific target groups of public policy vary between countries, overall public programmes focus on: the unemployed (especially the long-term unemployed), immigrants, low-skilled, low-income groups, rural or remote dwellers and people with disabilities.

Not surprisingly, most training undertaken by employed people is financed by the enterprises, while governments fund a large proportion of training for the unemployed and for those not in the labour force. More interesting is the observation that there is a positive relationship between the level of public support and the incidence of adult education for the low-skilled.

‘All countries have a broad variety of partners involved in adult learning, including ministries of education, ministries of labour, regional governments, local-level governments, educational institutions, special adult learning institutions and the social partners. Overall, there is no one good way to coordinate activities and policy-making, as

they depend on historical development and political, administrative and social frameworks. What counts is coordination across the board to attain policy coherence, and this may be best attained with the creation of specific institutions devoted to the endeavour’ (OECD, 2003: 87). Such specific institutions have been identified in Denmark, Portugal, Finland and the UK. In Norway, it is by a merger of different institutions that this issue has been addressed.

Table A11. Recommendations to enhance access and participation in adult learning: five key areas

Making learning more attractive to adults	Stimulate employment related learning	Developing financial incentives to invest in human capital of adults	Improve the quality of adult learning	Adopt a coordinated approach to adult learning to rationalise the use of resources and the learning landscape
<p>Pedagogical methods that are learner-centred, contextualised and relevant to their experiences adapted to adults.</p> <p>Flexible provision to suit adults' schedules, modularisation, ICT and distance education.</p> <p>Guidance and counselling to reach adults that otherwise would not consider or have little motivation to learn.</p> <p>Recognition of prior learning to avoid wasting time in relearning what is already known.</p>	<p>Overcome time and cost barriers through education and training leave.</p> <p>Access to skills assessment and skills development in firms for groups at risk.</p> <p>Flexible training in public employment services (modules, continuous admission and certification, etc.).</p> <p>Avoid sole quantitative results as financing criteria for unemployed training.</p>	<p>Introduction of individual incentives mechanisms (loans, grants, individual learning accounts).</p> <p>Offer entitlements for learning or study leave during working hours.</p> <p>Subsidy to private suppliers or individuals (compensate part of opportunity costs).</p> <p>Establish enterprise training levies or set up of national or sectoral training funds.</p>	<p>Introduction of quality insurance systems.</p> <p>Set standards for delivery and certify publicly the achievement of these standards.</p> <p>Introduce evaluation as an integral part of policy design and use more research and analysis as support for policy choice.</p>	<p>Rationalise and improve the coherence of supply and coordinate actors in the public adult learning system.</p> <p>Coordination between government and non-governmental actors (employers, trade unions, public and private providers, community groups).</p> <p>Balance between top-down (government defines structures and financing) and bottom-up (feedback from local actors and room for innovation).</p> <p>Promote partnerships and coordination across sectors and actors.</p>

ANNEX 5

Evaluation associations

Country	Name/Contact	Country	Name/Contact
Belgium	Société Wallonne de l'évaluation et de la prospective Institut Jules-Destrée 3, rue du Château B-6032 Charleroi E-mail: admin-reseaux@destree.org Internet: http://www.prospeval.org/	Ireland	Irish evaluation network Robert Holton The Policy Institute, Trinity College One College Green Dublin 2 – Ireland Internet: http://www.policyinstitute.tcd.ie/Irish_Evaluation_Network.html E-mail: evaluation.network@tcd.ie
Denmark	Dansk Evaluerings Selskab (Danish Evaluation Society), c/o AKF Kristine Bang Nielsen Nyropsgade, 37 DK-1602 København V E-mail: des@akf.dk Internet: http://www.danskevalueringsselskab.dk/	Italy	Associazione Italiana di Valutazione Bruno Vigilio Turra Piazzale Stazione, 7 I-35131 Padova c/o Segreteria AIV Internet: http://www.valutazioneitaliana.it E-mail: segreteria@valutazioneitaliana.it
Germany	Deutsche Gesellschaft für Evaluation e.V. (DeGEval) Bücheler Weg 27 D-53347 Alfter (bei Bonn) E-mail: info@degeval.de Internet: http://www.degeval.de/		Istituto Italiano di Valutazione Srl Sede legale: Via Mauro Macchi 27, I-20124 Milano E-mail: valutazione@email.it Internet: http://www.valutazione.it/iiv.htm
Spain	Spanish Public Policy Evaluation Society E-mail: carmenvelez@idr.es Internet: http://www.idr.es (with restricted access)	The Netherlands	VIDE – Beroepsvereniging voor toezichthouders, inspecteurs, handhavers en evaluatoren Chair: Frans L. Leeuw Secretariat: Heidi van Steeg Ambachtsstraat 15 Postbus 1058 3861 RH Nijkerk 3860 BB Nijkerk The Netherlands The Netherlands Internet: http://www.videnet.nl/ E-mail: secretariaat@videnet.nl
France	Société Française de l'Évaluation 1 Avenue de la porte de Montrouge F-Paris E-mail: evaluation.sfe@wanadoo.fr Internet: http://www.sfe.asso.fr	Austria	Plattform Forschungs- und Technologieevaluierung (Platform Research and Technology Policy Evaluation) Wiedner Hauptstrasse 76/I A-1040 Wien Internet: www.fteval.at E-mail: office@fteval.at
	Centre européen d'expertise en évaluation Secrétariat – administration: Françoise Revellin, Christine Gimenez 13bis, place Jules Ferry F-69006 Lyon Internet: http://www.c3e.fr/ E-mail: lyon@eureval-c3e.fr		

Country	Name/Contact	Country	Name/Contact
Austria	Austrian evaluators are also engaged in the German Evaluation Society (DeGEval) mentioned above.	Switzerland	Commission externe d'évaluation des politiques publiques (CEPP; External Commission for Public Policy Evaluation) rue du Stand 20 bis case postale 3937 CH-1211 Genève 3 Internet: http://www.geneve.ch/cepp/
Portugal	'The evaluation community in Portugal has taken the first firm steps in the direction of setting up a national Portuguese evaluation society.' Source: European Evaluation Society website, 21 April 2004.	Russia	International Program Evaluation Network Internet: http://ipen21.org/ipen/en/default.html
Finland	Finnish Evaluation Society Contact: Johanna Nurmi E-mail: johanna.nurmi@vm.vn.fi Internet: http://www.finnishevaluationsociety.net	Canada	Canadian Evaluation Society Secretariat: 1485 Laperriere Avenue Ottawa, Ontario K1Z 7S8 E-mail: secretariat@evaluationcanada.ca Internet: http://www.evaluationcanada.ca/
Sweden	Utvärderarna (Swedish Evaluation Society) c/o Statskontoret PO Box 2280 S-103 17 Stockholm Internet: http://www.statskontoret.se/utvarderarna/hem.html E-mail: jenny.soukkan@statskontoret.se	European Evaluation Society Ane Møller – Administrative Officer Political Science and Public Management University of Southern Denmark DK 5230 Odense M E-mail: ees@sam.sdu.dk Internet: http://www.europeanevaluation.org	
UK	UK Evaluation Society c/o Professional Briefings 120 Wilton Road London SW1V 1JZ United Kingdom Internet: http://www.evaluation.org.uk/ E-mail: ukes@profbriefings.co.uk	African Evaluation Association (AfrEA) PO Box 41829 Craighall 2024 Johannesburg South Africa E-mail: info@afrea.org Internet: http://www.afrea.org	
	The Tavistock Institute 30 Tabernacle Street London EC2A 4UE United Kingdom E-mail: central.admin@tavinstitute.org Internet: http://www.tavinstitute.org/	American Evaluation Association 16 Sconticut Neck Rd #290 USA-Fairhaven MA 02719 E-mail: AEA@eval.org Internet: http://www.eval.org/	
Poland	Polskie Towarzystwo Ewaluacyjne (Polish Evaluation Society) ul. Elektronowa 2 p. 14 PL-03219 Warszawa Internet: http://www.pte.org.pl/	International Organization for Cooperation in Evaluation (IOCE) Internet: http://home.wmis.net/~russon/ioce	
Switzerland	SEVAL (Swiss Evaluation Society) c/o SANW Bärenplatz 2 CH-3011 Bern E-mail: kissling@sanw.unibe.ch Internet: http://www.seval.ch/en/index.cfm	WWW Virtual Library: Evaluation The World Wide Evaluation Information Gateway Internet: http://www.policy-evaluation.org	

ANNEX 6

Journals on evaluation in Europe and beyond: alphabetical list ⁽⁶⁾

- (a) *The American journal of evaluation*
http://www.eval.org/american_journal_of_evaluation.htm
Articles deal with topics applicable to the broad field of programme evaluation. Articles may focus on evaluation methods, theory, practice or findings. Examples of contributions include, but are not limited to, reviews of new developments in evaluation, descriptions of a current evaluation study, critical reviews of some area of evaluation practice and presentations of important new techniques.
-
- (b) *Assessment*
ISSN 1073-1911
The scope of the journal extends from the evaluation of individuals and groups in clinical, counselling, health, forensic, organisational, industrial, and educational settings to the assessment of treatment efficacy, programme evaluation, job performance and the study of behaviour outcomes.
-
- (c) *Assessment in education: principles, policy and practice*
ISSN 0969-594X
<http://www.tandf.co.uk/journals/titles/0969594X.asp>
Recent decades have witnessed significant developments in educational assessment. New approaches to assessing student achievement have been complemented by the increasing prominence of educational assessment as a policy issue. In particular, there has been a growth of interest in modes of assessment that promote, as well as measure, standards and quality. These have profound implications for individual learners, institutions and the educational system itself.
-
- (d) *Berufsbildung in Wissenschaft und Praxis*
<http://www.bibb.de/de/360.htm>
BWP - *Berufsbildung in Wissenschaft und Praxis* [Vocational education in theory and practice] is a bimonthly journal published by the German Federal Institute of Vocational Training (BIBB) and informs about:
- (i) research results of BIBB projects;
 - (ii) new developments in vocational education and training;
 - (iii) training policy;
 - (iv) new and updated training regulations;
 - (v) training of trainers;
 - (vi) international aspects of VET;
 - (vii) new educational technologies;
 - (viii) professional standards and skills development, etc.
-
- (e) *Canadian journal of program evaluation*
ISSN 0834-1516
Full text on: http://www.evaluationcanada.ca/site.cgi?s=4&ss=2&_lang=en
Articles on all aspects of the theory and practice of evaluation, including methodology, evaluation standards, implementing evaluations, reporting and use of studies and the audit or metaevaluation of evaluation.
-
- (f) *Educational evaluation and policy analysis*
ISSN 0162-3737 – <http://www.aera.net/pubs/eepa/>
Educational evaluation and policy analysis (EEPA) publishes scholarly articles concerned with important issues in formulating, implementing, and evaluating education policy. EEPA is open to all of the diverse methodologies and theoretical orientations represented in AERA published work. We welcome submissions focused on international and comparative policy issues in education as well as domestic issues. Manuscripts should be written

⁽⁶⁾ The list of evaluation journals has been prepared by Cedefop's library and documentation services.

in a way that appeals to the broad and diverse interests of the EEPA readership, who work in a variety of institutional settings.

(g) *Educational research and evaluation*

ISSN 1380-3611 – <http://www.szp.swets.nl/szp/journals/er.htm>

Educational research and evaluation (ERE) is a journal of research relating to education practice. Its purpose is to provide an outlet for research from all nations and to communicate to the readers the findings of educational research from many perspectives, national contexts and methodologies. The journal places few limitations on content, focus or methods used in the articles, as long as they include disciplined inquiry into important issues of educational practice.

(h) *European journal of education*

<http://www.blackwellpublishing.com/journal.asp?ref=0141-8211&site=1>

The prime aims of the *European journal of education* are: to examine, compare and assess education policies, trends, reforms and programmes of European countries in an international perspective, to disseminate policy debates and research results to a wide audience of academics, researchers, practitioners and students of education sciences, to contribute to the policy debate at national and European level by providing European administrators and policy-makers in international organisations, national and local governments with comparative and up-to-date material centred on specific themes of common interest.

Special issue: Volume 38/2, *Reforming education and training: the role of evaluation*, guest editor: David Parkes.

(i) *Evaluation: the international journal of theory, research and practice*

<http://www.sagepub.co.uk/journal.aspx?pid=105547>

Over the last two decades, evaluation has become a major issue for academics, governmental and public organisations and businesses throughout the world. This has, however, resulted in a body of knowledge scattered across disciplines, professions and countries.

To promote dialogue internationally and to build bridges within this expanding field, *Evaluation: the*

international journal of theory, research and practice was launched in July 1995.

(j) *Evaluation and program planning*

ISSN 0149-7189

Evaluation and program planning is based on the principle that the techniques and methods of evaluation and planning transcend the boundaries of specific fields and that relevant contributions to these areas come from people representing many different positions, intellectual traditions and interests. In order to further the development of evaluation and planning, we publish articles from the private and public sectors in a wide range of areas: organisational development and behaviour, training, planning, human resource development, health and mental social services, mental retardation, corrections, substance abuse, and education. The primary goals of the journal are to assist evaluators and planners to improve the practice of their professions, to develop their skills and to improve their knowledge base.

(k) *Evaluation and research in education*

ISSN 0950-0790

<http://www.ingenta.com/journals/browse/mm/eri>

(l) *Evaluation review*

ISSN 0193-841X

<http://www.sagepub.com/journals/0193841X.htm>

For 26 years *Evaluation review* has served as a vital interdisciplinary forum for researchers, planners, and policy-makers who develop, implement, and utilise studies designed to improve the human condition. *Evaluation review* brings you the latest applied evaluation methods used in a wide range of disciplines, including: education, public health, criminal justice, child development, mental health, social work, public administration and environmental studies. *Evaluation review* keeps you up-to-date on the latest quantitative and qualitative methodological developments, as well as related applied research issues. Additional features include periodic review essays, research briefs of continuing or completed studies and craft reports on innovative applications of evaluation research techniques and concepts.

(m) *Journal of personnel evaluation in education*

ISSN 0920-525X – <http://www.kluweronline.com/issn/0920-525X/>

The *Journal of personnel evaluation in education* publishes research and applied scholarship perspectives on current issues in the evaluation or assessment of teachers, administrators, educational specialist positions, and higher education faculty performance. It provides a forum for prominent professionals to discuss, analyse and debate issues crucial to understanding, implementing and advancing effective personnel evaluation policies, programmes and practices. Priority consideration is given to manuscripts that advance the understanding and application of evaluation or assessment to inform teacher, school leader and educator about quality. Included under the rubric of quality are issues and concerns related, for example, to teacher and school leader preparation, licensure/certification, induction, performance appraisal, professional development and retention, as well as the exploration and impact of non-cognitive factors (dispositions, attitudes, beliefs) on the quality of instruction and learning. The common theme for any and all such investigations is the clear connection between the use of evaluation and assessment to understand and improve the quality of teaching, school leadership and/or students' educational experiences. The *Journal of personnel evaluation in education* welcomes diverse professional perspectives and experiences in the areas of evaluation and assessment in educational contexts. Researchers, policy-makers and practitioners are encouraged to submit manuscripts addressing current developments in the theory and application of education-based evaluation and assessment.

(n) *Journal of vocational education and training*

ISSN 1363-6820 – <http://www.triangle.co.uk/vae/>
Journal of vocational education and training is a fully-refereed international journal that publishes scholarly articles addressing the development of practice and theory in work-related education, wherever that education occurs. In many industrial countries education for occupations is mainly a post-compulsory provision, though that is fast disappearing as those societies realise the fruitful

potential for using work-oriented studies for general educational purposes in compulsory primary and secondary schooling. In developing countries, where schooling can only claim resources if it is more directly relevant to national economic needs, vocational preparation is more often a normal experience for school children. This journal seeks to encourage such developments by acting as a focal point for the study of all aspects of vocational and prevocational education throughout the world.

(o) *National VET research and evaluation program: a newsletter for researchers who work on national VET research and evaluation projects (NCVER – Australia)*

<http://www.ncver.edu.au/publications/1421.html>

A newsletter for researchers who work on national VET research and evaluation projects

(p) *New directions for evaluation*

ISSN 1097-6736

<http://www.josseybass.com/WileyCDA/WileyTitle/productCd-EV.html>

New directions for evaluation offers programme administrators, institutional researchers and all evaluation specialists techniques and procedures for conducting useful evaluations of all types of programmes, from educational curricula to health programmes.

(q) *Research evaluation*

<http://iris.ingentaselect.com/vl=5632385/cl=137/nw=1/rpsv/cw/beechn/09582029/contp1.htm>

Research evaluation is a peer-reviewed, international journal. It ranges from the individual research project up to inter-country comparisons of research performance. Research projects, researchers, research centres, and the types of research output are all relevant. It includes public and private sectors, natural and social sciences. The term 'evaluation' applies to all stages from priorities and proposals, through the monitoring of continuing projects and programmes, to the use of the results of research.

(r) *Studies in educational evaluation*

ISSN 0191-491X

<http://www.ingenta.com/journals/browse/els/0191491x?mode=direct>

Studies in educational evaluation publishes original reports of evaluation studies. Four types of article are published by the journal:

- empirical evaluation studies representing evaluation practice in educational systems around the world;
- theoretical reflections and empirical studies related to issues involved in the evaluation of educational programmes, educational institutions, educational personnel and student assessment;
- articles summarising the state-of-the-art concerning specific topics in evaluation in general or in a particular country or group of countries;
- book reviews and brief abstracts of evaluation studies.

(s) *Zeitschrift für Evaluation* (DE)

ISSN 1619-5515 – <http://www.zfev.de/>

The journal publishes scientific and practice-oriented articles on: theoretical foundations of evaluation studies; methodological concepts for evaluation; examples of application for planning and executing evaluations; presentation of organisations and provision of evaluation; possibilities and limitations in the use of evaluation results; topical information on various thematic fields; developments for quality assurance of evaluation.

References

- Aakvik, A.; Heckman, J. J.; Vytlačil, E. J. *Treatment effects for discrete outcomes when responses to treatment vary among observationally identical person: an application to Norwegian vocational rehabilitation programmes*. Cambridge: National Bureau of Economic Research, 2000 (NBER Technical working paper, 262).
- Accountability and evaluation glossary. In: *Nonprofit good practice guide*. Available from Internet: <http://www.nonprofitbasics.org/TopicAreaGlossary.aspx?ID=14> [cited 23.8.2004].
- Acemoglu, D.; Pischke, J-S. *Beyond Becker: training in imperfect labor markets*. Cambridge: National Bureau of Economic Research, 1998 (NBER Working paper, 6740).
- Acemoglu, D.; Pischke, J-S. The structure of wages and investment in general training. *Journal of Political Economy*, 1999, Vol. 107, No 3, p. 539.
- Albæk, E.; Rieper, O. Evaluation in Danish governance. In: Furubo, J-E.; Rist, R. C.; Sandahl, R. (eds) *International atlas of evaluation*. London: Transaction Publishers, 2002, p. 27-46.
- Aldrich, R. The apprentice in history. In: Ainley, P.; Rainbird, H. (eds) *Apprenticeship towards a new paradigm of learning*. London: Kogan Page, 1999.
- Alesina, A.; La Ferrara, E. Participation in heterogenous communities. *Quarterly Journal of Economics*, 2000a, Vol. 115, Issue 3, p. 847-904.
- Alesina, A.; La Ferrara, E. *Who trusts others?* London: Centre for Economic Policy Research – CEPR, 2000b (Discussion paper, 2646).
- Allmendinger, J., *Lebensverlauf und Sozialpolitik: die Ungleichheit von Mann und Frau und ihr öffentlicher Ertrag*. Frankfurt/Main: Campus, 1994.
- Amin, A.; Graham, S. The ordinary city. *Transactions of the Institute of British Geographers*, 1997, Vol. 22, No 4, p. 411-429.
- Andreß, H.J. Instabile Erwerbskarrieren und Mehrfacharbeitslosigkeit. Ein Vergleich mit der Problemgruppe der Langzeitarbeitslosen. *Mitteilungen aus der Arbeitsmarkt- und Berufsforschung*, 1989, Vol. 22, No 1, p. 17-32.
- Angrist, J.D.; Levy, V. *The effect of teen childbearing and single parenthood on childhood disabilities and progress in school*. Cambridge: National Bureau of Economic Research, 1996 (NBER Working paper, 5807).
- Antoninis, M.; Tsakoglou, P. Who benefits from public education in Greece? Evidence and policy implications. *Education Economics*, 2001, Vol. 9, No 2, p. 197-222.
- Argyris, C.; Schön, D. *Organisational learning II, theory, method and practice*. Reading, MA: Addison-Wesley, 1996.
- Arias, O.; McMahon, W. Dynamic rates of return to education in the US. *Economics of Education Review*, 2001, Vol. 20, Issue 2, p. 121-138.
- Arthur, J. Effects of human resource systems on manufacturing performance and turnover. *Academy of Management Journal*, 1994, Vol. 37, Issue 3, p. 670-687.
- Arum, R.; Shavit, Y. Secondary vocational education and the transition from school to work. *Sociology of Education*, 1995, Vol. 68 (2), p. 187-204.
- Ashenfelter, O.; Krueger, A. Estimates of the economic returns to schooling from a new sample of twins. *American Economic Review*, 1994, Vol. 84, p. 1157-1173.
- Ashenfelter, O.; Rouse, C. How convincing is the evidence linking education and income? *Journal of Labor Economics*, 1994, Vol. 12, No 1.

- Ashenfelter, O.; Rouse, C. *Schooling, intelligence and income in America: cracks in the bell curve*. Cambridge: National Bureau of Economic Research, 1999 (NBER Working paper, W6902).
- Asplund, R. *Public funding and returns on education*. Ely: PJB Associates, 2003 (New perspectives on learning – Briefing paper 29).
- Asplund, R.; Pereira, P.T. *Returns to human capital in Europe: a literature review*. Helsinki: Taloustieto Oy, 1999.
- Asteriou, D.; Agiomirgianakis, G.M. Human capital and economic growth – Time series evidence from Greece. *Journal of Policy Modeling*, 2001, Vol. 23, Issue 5, p. 481-489.
- Aucoin, P. Administrative reforms in public management: paradigms, principles, paradoxes and pendulums. *Governance*, 1990, Vol. 3, Issue 2, p. 115-137.
- Auer, P.; Kruppe, Th. Monitoring of labour market policy in EU Member States. In: Schmid, G. et al. (eds) *International handbook of labour market policy and evaluation*. Cheltenham: Edward Elgar, 1996, p. 899-922.
- Autor, D.H. Why do temporary help firms provide free general skills training? *Quarterly Journal of Economics*, 2001, Vol. 116, Issue 4, p. 1409-1448.
- Bainbridge, S. et al. *Learning for employment: second report on vocational education and training policy in Europe: executive summary*. Luxembourg: EUR-OP, 2003.
- Bainbridge, S. et al. *Learning for employment: second report on vocational education and training policy in Europe*. Luxembourg: EUR-OP, 2004 (Cedefop Reference series, 51).
- Baldwin, J.R.; Johnson, J. Business strategies in more- and less-innovative firms in Canada. *Research Policy*, 1996, Vol. 25, Issue 5, p. 785-804.
- Ballot, G. *Firms investment in human capital: sponsoring and effect on performance*. Paper submitted to the European conference on 'The future of work: challenges for the European Employment Strategy', held in Athens the 13 and 14 February, 2003.
- Ballot, G.; Fakhfakh, F.; Taymaz, E. Firms human capital, R&D and performance: a study on French and Swedish firms. *Journal of Labour Economics*, 2001, Vol. 8, No 4, p. 443-462.
- Balsan, D.; Hanchane, S.; Werquin, P. Mobilité professionnelle initiale: éducation et expérience sur le marché du travail. *Formation-Emploi*, 1996, No 46, p. 31-46.
- Barceinas-Paredes, F. et al. *Unemployment and returns to education in Europe*. Barcelona: Universitat Autònoma, 2000 (Working paper).
- Baron, S.; Field, S.J.; Schuller, T. *Social capital: critical perspectives*. Oxford: University Press, 2000.
- Barrett, A. Economic performance of education and training: costs and benefits. In: Descy, P.; Tessaring, M. (eds) *Training in Europe. Second report on vocational training research in Europe 2000: background report*. Luxembourg: EUR-OP, 2001, Vol. II, Part 4, p. 383-404 (Cedefop Reference series).
- Barrett, A. et al. *Exploring the returns to continuing vocational training in enterprises; a review of research within and outside of the European Union*. Luxembourg: EUR-OP, 1998 (Cedefop Panorama series, 83).
- Barrett, A.; O'Connell, P.J. *Does training generally work? The returns to in-company training*. Bonn: IZA – Institute for the Study of Labour, 1999 (Discussion paper, 51).
- Barro, R.J. Economic growth in a cross section of countries. *Quarterly Journal of Economics*, 1991, Vol. 106, p. 407-443.
- Barro, R.J. *Determinants of economic growth: a cross-country empirical study*. Cambridge, MA: MIT Press, 1997.
- Barro, R.J. *Education and economic growth*. Paper presented at the international symposium on the contribution of human and social capital to sustained economic growth and well-being, organised by the OECD and HRDC, Quebec City, Canada, 19-21 March 2000 (Mimeo).

- Barro, R.J. Human capital and growth. *American Economic Review*, 2001, Vol. 91, No 2, p. 12-17.
- Barro, R.J.; Lee, J.W. Sources of economic growth. *Carnegie-Rochester Conference Series on Public Policy*, 1994, Vol. 40, p. 1-46.
- Barro, R.J.; Lee, J.W. International measures of schooling years and schooling quality. *American Economic Review*, 1996, Vol. 86, No 2, p. 218-223 (Papers and proceedings).
- Barron, J.M. et al. *On-the-job training*. Kalamazoo, MI: Upjohn Institute for Employment Research, 1997.
- Barron, J.M.; Berger, M.C.; Black, D.A. Do workers pay for on-the-job training? *Journal of Human Resources*, 1999, Vol. 34, No 2, p. 235-252.
- Bartel, A.P. *Productivity gains from the implementation of employee training programs*. Cambridge: National Bureau of Economics Research, 1991 (NBER Working paper, 3893).
- Bartel, A.P. Productivity gains from the implementation of employee training programs. *Industrial Relations*, 1994, Vol. 33, p. 411-425.
- Bartel, A.P. Training, wage growth and job performance: evidence from a company database. *Journal of Labor Economics*, 1995, Vol. 13, No 3, p. 401-425.
- Bassanini, A.; Scarpetta, S. *Does human capital matter for growth in OECD countries? Evidence for people mean-group estimates*. Paris: OECD, 2001 (Working paper, 282).
- Bassegy, M. *Case study research in educational settings*. Maidenhead: Open University Press, 2003.
- Bassi, L.J. et al. *Human capital investments and firm performance*. Washington: Human Capital Dynamics, 2001 (Working paper).
- Bassi, L.J. et al. Profiting from learning: firm level effects of training investments and market implications. *Singapore Management Review*, 2002, Vol. 24, No 3, p. 61-76.
- Baumgartl, B.; Strietska-Ilina, O.; Schaumberger, G. Consultancy for free? Evaluation practice and culture in the European Union and eastern Europe. Findings from selected EU programmes. In: Descy, P.; Tessaring, M. (eds) *Evaluation of systems and programmes. Third report on vocational training research in Europe: background report*. Luxembourg: EUR-OP, 2004 (Cedefop Reference series, 57).
- Baumol, W.J.; Blackman, S.A.B.; Wolff, E.N. *Productivity and American leadership: the long view*. Cambridge, Mass.: MIT Press, 1989.
- Beck, K. (ed.) *Evaluation and assessment of flexibility, mobility and transferability in European countries*. Proceedings of COST A11: working group 5, working document No 1. European Commission, Directorate-General for Research. Luxembourg: EUR-OP, 2000.
- Beck, U. *Risk society. Towards a new modernity*. London: SAGE, 1994.
- Becker, G.S. Investment in human capital: a theoretical analysis. *Journal of Political Economy*, 1962, Vol. 70, No 5, p. 9-49.
- Becker, R. Bildung und Lebenserwartung in Deutschland. Eine empirische Längsschnittuntersuchung aus der Lebensverlaufsperspektive. *Zeitschrift für Soziologie*, 1998, Vol. 27, No 2, p. 133-150.
- Becker, R.; Schömann, K. Berufliche Weiterbildung und Einkommensdynamik: eine Längsschnittstudie mit besonderer Berücksichtigung von Selektionsprozessen. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 1996, Vol. 48, No 3, p. 426-461.
- Becker, R.; Schömann, K. Berufliche Weiterbildung und Einkommenschancen im Lebensverlauf: Empirische Befunde für Frauen und Männer in West- und Ostdeutschland. In: Beer, D. et al. (eds) *Die wirtschaftlichen Folgen der Aus- und Weiterbildung*. Munich: Hampp, 1999.
- Beer, D. et al. (eds) *Die wirtschaftlichen Folgen der Aus- und Weiterbildung*. Munich: Hampp, 1999.
- Bellmann, L. Vocational training research on the basis of enterprise surveys: an international perspective. In: Descy, P.; Tessaring, M. (eds) *Training in Europe: second report on vocational training*

- research in Europe 2000. Background report.* Luxembourg: EUR-OP, 2001, Vol. 2, p. 279-312 (Cedefop Reference series).
- Bellmann, L.; Büchel, F. *Betrieblich finanzierte Weiterbildung und Unternehmenserfolg: eine Analyse für West- und Ostdeutschland unter besonderer Berücksichtigung von Selektionseffekten.* Nürnberg: Institut für Arbeitsmarkt- und Berufsforschung – IAB, 2000 (Working paper).
- Bemelmans-Videc, M.L. Evaluation in the Netherlands 1990-2000: consolidation and expansion. In: Furubo, J.E.; Rist, R.C.; Sandahl, R. (eds) *International atlas of evaluation.* London: Transaction Publishers, 2002, p. 93-114.
- Bender, S.; Dietrich, H. *Unterschiedliche Startbedingungen haben langfristige Folgen: der Einmündungsverlauf der Geburtskohorten 1964 und 1971 in Ausbildung und Beschäftigung – Befunde aus einem IAB Projekt.* Nuremberg: IAB, 2001 (Serie Werkstattbericht, 11).
- Benhabib, J.; Spiegel, M.M. The role of human capital in economic development: evidence from aggregate cross-country data. *Journal of Monetary Economics*, 1994, Vol. 34, Issue 2, p. 143-173.
- Bergemann, A. et al. *Multiple active labor market policy participation in East Germany: an assessment of outcomes.* Halle: Institute of Economic Research, 2000 (Working paper).
- Berger-Schmitt, R. *Social cohesion as an aspect of the quality of societies: concept and measurement.* Mannheim: 2000 (EuReporting working paper, 14).
- Bewley, T.F. Why not cut pay? An empirical investigation in northern Nigeria. *European Economic Review* 1998, Vol. 42, Issue 3, p. 459-490.
- Beywl, W.; Geiter, C. *Evaluation. Controlling. Qualitätsmanagement in der betrieblichen Weiterbildung. Kommentierte Auswahlbibliographie.* Revised edition. Bielefeld: Bertelsmann, 1996.
- Beywl, W.; Speer, S. Developing standards to evaluate vocational education and training programmes. In: Descy, P.; Tessaring, M. (eds) *The foundations of evaluation and impact research. Third report on vocational training research in Europe: background report.* Luxembourg: EUR-OP, 2004 (Cedefop Reference series, 58).
- Bîrzea, C. The dilemma of the reform of Romanian education: shock therapy, the infusion of innovation or cultural decommunisation? *Higher Education in Europe*, 1997, Vol. 22, No 3, p. 321-328.
- Bishop, J. The impact of previous training on productivity and wages. In: Lynch, L.M. (ed.) *Training and the private sector: international comparisons.* Chicago: University press, 1994.
- Björklund, A.; Regnér, H. Experimental evaluation of European labour market policy. In: Schmid, G.; O'Reilly, J.; Schömann, K. (eds) *International handbook of labour market policy and evaluation.* Brookfield: Edward Elgar, 1996, p. 89-114.
- Bjørnåvold, J. *Making learning visible: identification, assessment and recognition of non-formal learning in Europe.* Luxembourg: EUR-OP, 2000 (Cedefop Reference series).
- Black, S.; Lynch, L. Human capital investments and productivity. *American Economic Review*, 1996, Vol. 86, No 2, p. 263-267.
- Black, S.; Lynch, L. *How to compete: the impact of work practices and information technology on productivity.* Cambridge: National Bureau of Economic Research, 1997 (NBER Working paper, w6120).
- Blackwell, L.; Bynner, J. *Learning, family formation and dissolution.* London: Centre for Research on the Wider Benefits of Learning, Institute of Education, 2002.
- Blandy, R. et al. *Does training pay? Evidence from Australian enterprises.* Adelaide: NCVET – National Centre for Vocational Education Research, 2000.
- Blaug, M. *The methodology of economics, or how economists explain.* Cambridge: University Press, 1992.
- Blechinger, D.; Pfeiffer, F. *Qualification, employment and technical progress: further empirical evidence based on data of the Mannheim innovation panel.* Mannheim:

- ZEW – Centre for European Economic Research, 1998 (Discussion paper, 98-04).
- Bloom, H. et al. The benefits and costs of JTPA Title II-A programs: key findings from the National Job Training Partnership Act study. *Journal of Human Resources*, 1997, Vol. 32, No 3, p. 549-576.
- Bloomer, M.; Hodkinson, P. Learning careers: continuity and change in young people's dispositions to learning. *British Educational Research Journal*, 2000, Vol. 26, No 5, p. 583-596.
- Blossfeld, H.P. *Kohortendifferenzierung im Karriereprozess. Eine Längsschnittstudie über die Veränderung der Bildungs- und Berufschancen im Lebenslauf*. Frankfurt am Main: Campus, 1989.
- Blossfeld, H.P. *Bildung, Arbeit und soziale Ungleichheit im Globalisierungsprozess*. Bielefeld: Faculty of Sociology, 2000 (Globalife working paper series, 1).
- Blossfeld, H.P.; Huinink, J. Bildung, Karriere und das Alter bei der Eheschließung von Frauen. In: Felderer, B. (ed.) *Bevölkerung und Wirtschaft*. Berlin: Duncker & Humblot, 1990, p. 539-553.
- Blossfeld, H.P.; Mayer, K.U. Berufsstruktureller Wandel und soziale Ungleichheit. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 1998, Vol. 43, No 4, p. 671-696.
- Blossfeld, H.P.; Timm, A. *Das Bildungssystem als Heiratsmarkt. Eine Längsschnittanalyse der Wahl von Heiratspartnern im Lebenslauf*. Bremen: Sonderforschungsbereich 186 der Universität Bremen, 1997 (Sfb-Arbeitspapier, 43).
- Blundell, R. et al. Human capital investment: the returns from education and training to the individual, the firm and the economy. *Fiscal Studies*, 1999, Vol. 20, No 1, p. 1-23.
- Blundell, R.; Dearden, L.; Sianesi, B. *Estimating the returns to education: models, methods and results*. London: Centre for the Economics of Education, London School of Economics, 2001 (Discussion paper 16).
- Bonnal, L.; Fougère, D.; Sérandon, A. L'impact des dispositifs d'emploi sur le devenir des jeunes chômeurs: une évaluation économétrique sur données longitudinales, *Economie et Prévision*, 1994, No 115, p. 1-28.
- Bonnal, L.; Fougère, D.; Sérandon, A. Evaluating the impact of French employment policies on individual labour market histories. *Review of Economic Studies*, 1997, Vol. 64, Issue 4, p. 683-713.
- Boockmann, B.; Steiner, V. *Cohort effects and the return to education in West Germany*. Mannheim: ZEW – Centre for European Economic Research, 2000 (Discussion paper, 00-05).
- Booth, A.L.; Zoega, G. *Is wage compression a necessary condition for firm-financed general training? A comment on Acemoglu and Pischke*. Colchester: ILR – University of Essex, 1999 (Working paper).
- Born, C. Erstausbildung und weiblicher Lebenslauf. *Zeitschrift für Soziologie der Erziehung und Sozialisation*, 2000, No 3 (Beiheft), p. 50-65.
- Bosma, N. et al. *The value of human and social capital investments for the business performance of start-ups*. Amsterdam: Tinbergen Institute, 2002 (Discussion paper, TI 02-027/3).
- Bosworth, D.L. *Management skills in the UK*. Research report prepared for the Department for Education and Employment. Manchester: Manchester School of Management, UMIST, 1999.
- Böttger, A. *Gewalt und Biographie. Eine qualitative Analyse rekonstruierter Lebensgeschichten von 100 Jugendlichen*. Baden-Baden: Nomos, 1998.
- Boyd, R. Confirmation, semantics and the interpretation of scientific theories. In: Boyd, R.; Gasper, P.; Trout, J.D. *The philosophy of science*. Cambridge, MA: MIT Press, 1991.
- Brandsma, J. (ed.) *The effectiveness of labour market oriented training for the long-term unemployed. Final report of the TSER project*. Brussels: European Commission, 1999.
- Bratberg, E.; Nilsen, O.A. *Transition from school to work: search time and job duration*. Bonn: IZA – Institute for the Study of Labor, 1998 (Discussion paper, 27).

- Brauns, H.; Müller, W.; Steinmann, S. *Educational expansion and returns to education: a comparative study on Germany, France, the UK and Hungary*. Mannheim: MZES – Mannheimer Zentrum für Europäische Sozialforschung, 1997 (Working paper I, 23).
- Breen, R. *The persistence of class origin inequalities among school leavers in the Republic of Ireland, 1984-1993*. Paper presented to the ESF workshop on transitions in youth: comparisons over time and across countries, held in Oostvoorne, the Netherlands, 22-25 September, 1995.
- Brinkmann, C. Zur Evaluation der aktiven Arbeitsförderung nach dem SGB III – Monitoring und Wirkungsforschung im Umbruch. *Mitteilungen aus der Arbeitsmarkt- und Berufsforschung*, 2000, Vol. 3, p. 483-499.
- Broadberry, S.N.; Wagner, K. Human capital and productivity in manufacturing during the twentieth century: Britain, Germany and the United States. In: Van Ark, B.; Crafts, N. (eds) *Quantitative aspects of post-war European economic growth*. Cambridge: University Press, 1996.
- Brodaty, T.; Crepon, B.; Fougère, D. Using matching estimators to evaluate alternative youth employment programmes: evidence from France, 1986-1988. In: Lechner, M.; Pfeiffer, F. (eds) *Econometric evaluation of labour market policies*. Heidelberg: Physika Verlag, 2001 (ZEW Economic Studies, Vol. 13).
- Brown, P.; Green, A.; Lauder, H. *High skills: globalisation, competitiveness and skill formation*. Oxford: University Press, 2001.
- Brunello, G. *On the complementarity between education and training in Europe*. Bonn: IZA – Institute for the Study of Labor, 2001a (Discussion paper, 309).
- Brunello, G. *Unemployment, education and earnings growth*. PURE project, TSER Programme. Brussels: European Commission, 2001b. Available from Internet: www.etla.fi/PURE/publications.htm [cited 27.8.2004].
- Brunello, G.; Comi, S. *Education and earnings growth: evidence from 11 European countries*. PURE project, 2000. Available from Internet: <http://www.etla.fi/PURE/Italy%20-%20Egrowth52.pdf> [cited 27.8.04].
- Brunello, G.; Comi, S.; Lucifora, C. Education, earnings growth and cohorts. In: PURE. *Public funding and private returns to education. Final Report*. 2001b, p. 22-34 2001 (Project Final report series – TSER Area 2 – SOE2-CT98-2044).
- Brunello, G.; Comi, S.; Lucifora, C. *The returns to education in Italy. A new look at the evidence*. Bonn: IZA – Institute for the Study of Labor, 2001a (Discussion paper, 130).
- Brunello, G.; Miniaci, R. The economic returns to schooling for Italian men. An evaluation based on instrumental variables. *Labour Economics*, 1999, No 6, Issue 4, p. 509-519.
- Büchel, F. *Die Qualität der Wiederbeschäftigung nach ununterbrochener und nach 'perforierter' Langzeitarbeitslosigkeit*. Nürnberg: IAB, 1992 (BeitrAB 162).
- Büchel, F. *The effects of overeducation on productivity in Germany: the firms' viewpoint*. Bonn: IZA – Institute for the Study of Labor, 2000 (Discussion paper, 216).
- Bukodi, E.; Robert, P. *The effect of the globalization process on the transition to adulthood in Hungary*. Bamberg: Department of Sociology, Bamberg University, 2002 (Globalife working paper series, 27).
- Buschor, E. Evaluation and new public management. *Zeitschrift für Evaluation*, 2002, Vol. 1, p. 61-73.
- Bynner, J. *Changing needs, new structures? Lessons for vocational preparation*. CEDAR International Conference, University of Warwick, 15 April 1994.
- Bynner, J. Skills and disadvantage in labour-market entry. In: Haubrich, K.; Reubel, F. (eds) *Berufsstart benachteiligter Jugendlicher in Europa. Konzepte zur beruflichen Integration im regionalen Kontext*. München: DJI, 1995 (Arbeitspapier, 2).
- Bynner, J. Education for what? *Education and Training*, 1998, Vol. 40, No 1, p. 4-5.
- Bynner, J.; et al. *Improving adult basic skills: benefits to the individual and to society*. London: HMSO, 2001 (DfEE Research report RR251).

- Bynner, J.; Parsons, S. Qualifications, basic skills and accelerating social exclusion. *Journal of Education and Work*, 2001, Vol. 14, Issue 3, p. 279-291.
- Calmfors, L. Active labour market policy and unemployment: a framework for the analysis of crucial design features. *OECD Economic Studies*, 1994, Vol. 22, p. 7-47.
- Calmfors, L.; Forslund, A.; Hemström, M. *Does active labour market policy work? Lessons from Swedish experience*. Uppsala: IFAU – Institute for labour market policy evaluation, 2002 (Working paper series, 4).
- Caselli, F.; Esquivel, G.; Lefort, F. Reopening the convergence debate: a new look at cross-country growth empirics. *Journal of Economic Growth*, 1996, Vol. 1, Issue 3, p. 363-389.
- Cedefop. *Lifelong learning: citizens' views*. Luxembourg: EUR-OP, 2003.
- Chelimsky, E. Politics, policy and research synthesis. *Evaluation: the International Journal of Theory, Research and Practice*, 1995, Vol. 1, No 1, p. 97-104.
- Chen, H.-T. *Theory-driven evaluation*. London: Sage, 1990.
- Chevalier, A. et al. The returns to education in the UK. In: Asplund, R.; Pereira, P.T. (eds) *Returns to human capital in Europe: a literature review*. Helsinki: Taloustieto Oy, 1999.
- Cockx, B.; Bardoulat, I. *Vocational training: does it speed up the transition rate out of unemployment?* Amsterdam: Tinbergen Institute, 2000 (Discussion paper, 9932).
- Cockx, B.; Van der Linden, B.; Karaa, A. A active labour market policy and job tenure. *Oxford Economic Papers*, 1998, Vol. 50, p. 685-708.
- Coffield, F. The holes in the heart of current policies on lifelong learning. In: *Policy, practice and partnership: getting to work on lifelong learning*. Report on the Conference held at Cedefop, Thessaloniki, Greece, 2-3 June 2003, p. 40-52.
- Cohen, D.; Soto, M. *Growth and human capital: good data, good results*. London: CEPR – Centre for Economic Policy Research, 2001 (Discussion paper series, 3025).
- Cohen, L.; Manion, L. *Research methods in education*. Fourth edition. London: Routledge, 1994.
- Cohen, L.; Manion, L.; Morrison, K. *Research methods in education*. Fifth edition. London: Routledge Falmer, 2000.
- Coleman, J. Social capital in the creation of human capital. *American Journal of Sociology*, 1988, Vol. 94, p. 95-120.
- Coleman, J. *Foundations of social theory*. Cambridge, Mass: Harvard University Press, 1990.
- Coles, B. et al. *Literature review of the costs of being 'not in education, employment or training' at age 16-18*. London: DfEE – Department for Employment and Education, 2002 (Research brief, 347).
- Coles, M. Evaluating the impact of reforms of vocational education and training: examples of practice. In: Descy, P.; Tessaring, M. (eds) *Evaluation of systems and programmes. Third report on vocational training research in Europe: background report*. Luxembourg: EUR-OP, 2004 (Cedefop Reference series, 57).
- Cook, T.D.; Campbell, D.T. *Quasi-experimentation: design and analysis issues for field settings*. Boston, MA: Houghton Mifflin Co, 1979.
- Council of the European Union. Council decision of 6 December 1994 establishing an action programme for the implementation of a European Community vocational training policy (1994/819/EC). *Official Journal of the European Communities*, 29 December 1994, L340, p. 8-24.
- Council of the European Union. Council decision of 26 April 1999 establishing the second phase of the Community vocational training action programme 'Leonardo da Vinci' (1999/382/EC). *Official Journal of the European Communities*, 1999a, 11 June 1999, L146, p. 33-47.
- Council of the European Union. Detailed work programme on the follow-up of the objectives of education and training systems in Europe. *Official Journal of the European Communities*, 14 June 2002, C142, p. 1-22.

- Council of the European Union. Council decision of 22 July 2003 on guidelines for the employment policies of the Member States. *Official Journal of the European Communities*, 5 August 2003, L197, p. 13-21.
- Crouch, C.; Finegold, D.; Sako, M. *Are skills the answer? The political economy of skill creation in advanced industrial countries*. Oxford: University Press, 1999.
- d'Arcimoles, C-H. Human resource policies and company performance: a quantitative approach using longitudinal data. *Organisation Studies*, 1997, Vol. 18, Issue 5, p. 857-874.
- Dahler-Larsen, P. *Evaluering og magt [Evaluation and power]*. Aarhus: University Press, 2004.
- Dahrendorf, R. et al. *Report on wealth creation and social cohesion in a free society*. London: Commission on Wealth Creation and Social Cohesion, 1995.
- Dale, A.; Egerton, M. *Highly educated women: evidence from the national child development study*. London: DfEE – Department for Employment and Education, 1997 (DfEE Research Studies, RS25).
- Danish Ministry of Finance. *Struktur- overvågning – International benchmarking af Danmark. [Structural monitoring – international benchmarking of Denmark]*. Copenhagen: Ministry of Finance, May 1999.
- Danish Ministry of Labour. *Effects of Danish employability enhancement programmes*. Copenhagen: Ministry of Labour, 2000 (Working group paper).
- Dar, A.; Tzannatos, Z. *Active labor market programmes: a review of the evidence from evaluations*. Washington DC: The World Bank, 1999 (Social protection discussion paper series, 9901).
- de la Fuente, A. *Public policies and private incentives to invest in education*. Paper submitted to the European conference on 'The future of work: key challenges for the European Employment Strategy', Athens, 13-14 February, 2003.
- de la Fuente, A.; Ciccone, A. *Human capital in a global and knowledge-based economy*. Final report. 2002.
- de la Fuente, A.; Doménech, R. *Human capital in growth regressions: how much difference does data quality make?* Paris: OECD, 2000 (ECO/WKP(2000)35).
- Dearden, L.; Reed, H.; van Reenen, J. *Who gains when workers train? Training and corporate productivity in a panel of British industries*. London: IFS – Institute for Fiscal Studies, 2000 (Working paper, w00/04).
- Deding, M.; Dall Schmidt, T. *Differences in income inequality across Europe – market driven or ...?* Colchester: EPAG – European Panel Analysis Group, 2002 (Working paper, 37).
- Delaney, J.T.; Huselid, M.A. The impact of human resource management practices on perceptions of organisational performance. *Academy of Management Journal*, 1996, Vol. 39, p. 949-969.
- Denny, K.; Harmon, C.; Lydon, R. *Cross-country evidence in the returns to education: patterns and explanations*. Mimeo, 2001.
- Denzin, N.K. *Interpretive ethnography: Ethnographic practices for the 21st century*. Thousand Oaks, CA: Sage, 1997.
- Denzin, N.K.; Lincoln, Y.S. (eds) *Handbook of qualitative research. Second edition*. Thousand Oaks, CA: Sage Publications, 2000.
- Derlien, H.U. Program evaluation in the Federal Republic of Germany. In: Rist, R. (ed.) *Program evaluation and the management of government: patterns and prospects across eight nations*. London: Transaction Publishers, 1990.
- Derlien, H.U. Policy Evaluation in Germany: institutional continuation and sectoral activation. In: Furubo, J.E.; Rist, R.C.; Sandahl, R. (eds) *International atlas of evaluation*. London: Transaction Publishers, 2002, p. 77-92.
- Derlien, H.U.; Rist, R.C. Policy evaluation in international comparison. In: Furubo, J.E.; Rist R.C.; Sandahl, R. (eds) *International atlas of evaluation*. London: Transaction Publishers, 2002, p. 139-456.

- Descy, P.; Tessaring, M. *Training and learning for competence. Second report on vocational education and training research in Europe: synthesis report*. Luxembourg: EUR-OP, 2001a (Cedefop Reference series, 6).
- Descy, P.; Tessaring, M. (eds) *Training in Europe. Second report on vocational training in Europe 2000: background report*. Luxembourg: EUR-OP, 2001b (Cedefop Reference series, 3008).
- Descy, P.; Tessaring, M. (eds) *Impact of education and training. Third report on vocational training research in Europe: background report*. Luxembourg: EUR-OP, 2004a (Cedefop Reference series, 54).
- Descy, P.; Tessaring, M. (eds) *The foundations of evaluation and impact research. Third report on vocational training research in Europe: background report*. Luxembourg: EUR-OP, 2004b (Cedefop Reference series, 58).
- Descy, P.; Tessaring, M. (eds) *Evaluation of systems and programmes. Third report on vocational training research in Europe: background report*. Luxembourg: EUR-OP, 2004c (Cedefop Reference series, 57).
- Dietz, G.-H.; Matt, E. Begrenzte Handlungsspielräume bei der Berufsfindung. Der Übergang von Haupt- und Sonderschülern in das Berufsbildungssystem. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 1994, No 5, p. 510-524.
- Diewald, M.; Sørensen, A. Erwerbsverläufe und soziale Mobilität von Frauen und Männern in West- und Ostdeutschland: makrostrukturelle Umbrüche und Kontinuitäten im Lebensverlauf. In: Diewald, M.; Mayer, K.U. (eds) *Zwischenbilanz der Wiedervereinigung*. Leverkusen: Leske und Budrich, 1996.
- Dolton, P.; O'Neill, D. Unemployment duration and the restart effect: some experimental evidence. *Economic Journal*, 1996, Vol. 106, p. 387-400.
- Doménech, R. *Human capital in growth regressions: how much difference does data quality make?* Mimeo, Universitat Autònoma de Barcelona, January 2000.
- Doucoulagos, C.; Sgro, P. *Enterprise return on a training investment*. Adelaide: NCVET – National Centre for Vocational Education Research, 2000.
- Drosdowski, G. (ed.) *Duden Etymologie. Herkunftswörterbuch der deutschen Sprache*. Second edition. Mannheim: Dudenverlag, 1989.
- Dunford, M. Divergence, instability and exclusion: regional dynamics in Great Britain. In: Lee, R.; Wills, J. (eds) *Geographies of economies*. London: Hodder Arnold, 1997, p. 259-277.
- Dunning, E. Towards a sociological understanding of football hooliganism as a world phenomenon. *European Journal on Criminal Policy and Research*. Dordrecht, Kluwer, June 2000, Vol. 8, No 2, p. 141-162.
- Easton, P.A. Sharpening our tools – Improving evaluation in adult and nonformal education. Hamburg: Unesco, 1996.
- Edquist, C.; Johnson, B. Institutions and organisations in systems of innovation. In: Edquist, C. (ed.) *Systems of innovation; technologies, institutions and organisations*. London: Pinter, 1997.
- Eisner, M. Modernisation, self-control and lethal violence: the long-term dynamics of European homicide rates in theoretical perspective. *British Journal of Criminology*, 2001, Vol. 41, No 4, p. 618-638.
- Englander, A.S.; Gurney, A. Medium-term determinants of OECD productivity. *OECD Economic Studies*, 1994, No 22, p. 49-109.
- Eraut, M. Do continuing professional development models promote one-dimensional learning? *Medical Education*, 2001a, No 35, p. 8-11.
- Eraut, M. Learning challenges for knowledge-based organisations. In: *Workplace learning in Europe*. London: CIPD – Chartered Institute of Personnel and Development, 2001b, p. 20-34.
- Eraut, M. The role and use of vocational qualifications. *National Institute Economic Review*, 2001c, No 178, p. 88-98.
- Erikson, R.; Jonsson, J. The Swedish context: educational reform and long-term change in educational inequality. In: Erikson, R.; Jonsson, J. (eds) *Can education be equalised? The Swedish test case in comparative perspective*. Boulder, CO: Westview Press, 1996a.

- Erikson, R.; Jonsson, J. Introduction: explaining class inequality in education: the Swedish test case. In: Erikson, R.; Jonsson, J. (eds) *Can education be equalised? The Swedish test case in comparative perspective*. Boulder, CO: Westview Press, 1996b.
- Ermisch, J.; Francesconi, M. Educational choices, families and young people's earnings. *Journal of Human Resources*, 2000, Vol. 35, No 1, p. 143-176.
- ETGACE. *Becoming active citizens: Europeans reflecting on their practice*. ETGACE – Education and training for governance and active citizenship in Europe, 2002 (Working paper).
- European Commission. *Evaluating socio-economic programmes*. Luxembourg: EUR-OP, 1999, Vol. 1–6 (MEANS collection).
- European Commission. *Final report on the implementation of the first phase of the community action programme Leonardo da Vinci (1995-1999): report from the Commission*. Luxembourg: EUR-OP, 2000 (COM(2000) 863 final).
- European Commission. *The concrete future objectives of the education systems*. Report from the Commission. Luxembourg: EUR-OP, 2001a (COM(2001) 59 final).
- European Commission, DG Employment and Social Affairs. *Conclusions of the ESF final evaluations*. Brussels: European Commission, 2001b.
- European Commission. *The social situation in the European Union 2002*. Luxembourg: EUR-OP, 2002a.
- European Commission. *Communication from the Commission – European benchmarks in education and training: follow-up to the Lisbon European Council*. Luxembourg: EUR-OP, 2002b (COM(2002) 629 final).
- European Commission. *Communication from the Commission to the Council, the European Parliament, the Economic and Social Committee and the Committee of Regions – taking stock of five years of the European employment strategy*. Luxembourg: EUR-OP, 2002c (COM(2002) 416 final).
- European Commission. *Report from the Commission – Interim report on the start of the operational implementation of the second phase of the Leonardo da Vinci programme (2000-2006)*. Luxembourg: EUR-OP, 2002d (COM(2002) 315 final).
- European Commission. *Communication from the Commission on impact assessment*. Luxembourg: EUR-OP, 2002e (COM(2002) 276 final).
- European Commission. DG Education and Culture. *Valorisation strategy for the Leonardo da Vinci vocational training programme*. Brussels: European Commission, 2002f (B/3/DS/ds D(2002) 1484, EC, February 2002).
- European Commission. *The future of the European Employment Strategy (EES): a strategy for full employment and better jobs for all*. Luxembourg: EUR-OP, 2003a (COM(2003) 6 final).
- European Commission, *The social situation in the European Union 2003*. Luxembourg: EUR-OP, 2003b.
- European Commission. Document de travail des services de la Commission: *mise en oeuvre du processus 'Education et formation 2010'*. Document d'appui pour le rapport intermédiaire conjoint sur la mise en oeuvre du programme de travail détaillé concernant le suivi des objectifs des systèmes d'éducation et de formation. Brussels: European Commission, 2003c (SEC).
- Eurostat. *Demographic consequences for the EU of the accession of twelve candidate countries*. Luxembourg: EUR-OP, 2001a (Statistics in focus: population and social conditions, 12/2001).
- Eurostat. *The EC household panel 'Newsletter' (1/01)*. Luxembourg: EUR-OP, 2001b (Statistics in focus: population and social conditions, 14/2001).
- Eurostat. *Public expenditure on labour market policies in 1999 varied greatly among Member States*. Luxembourg: EUR-OP, 2002 (Statistics in focus: population and social conditions, 12/2002).
- Eurostat. *The EC household panel 'Newsletter' (1/02)*. Luxembourg: EUR-OP, 2003

- (Method and nomenclatures: population and social conditions).
- Evans, K. *Shaping futures: learning for competence and citizenship*. Aldershot: Ashgate, 1998.
- Evans, K.; Heinz, W. *Becoming adults in England and Germany*. Rochester: Anglo-German Foundation, 1994.
- Fabrizio, B. *Educational performance and educational returns at entry into the Italian labour market*. Bamberg: Department of Sociology, 2000 (Globalife working paper series, 10).
- Fahrenkrog, G.; Delgado, L. *The societal bill: financing social protection and a societal environment*. Seville: Joint Research Center, 1999 (Future project series, 16).
- Farvaque, N.; Salais, R. *Implementing allowances for young people in difficulty in France: enhancing capabilities or increasing selectivity?* Ninth International Congress, Geneva 12-14 September. Geneva: BIEN, 2002. Available from Internet: <http://www.etes.ucl.ac.be/bien/Files/Papers/2002FarvaqueSalais.pdf> [cited 31.8.2004].
- Fay, R.G. *Enhancing the effectiveness of active labour market policies: evidence from programme evaluations in OECD countries*. Paris: OECD, 1996 (Labour market and social policy occasional papers, 18).
- Featherman, D.L.; Hauser, R.M. *Opportunity and change*. New York: Academic Press, 1978.
- Feldman, J. et al. National trends in educational differences in mortality. *American Journal of Epidemiology*, 1989, No 129, p. 919-933.
- Felstead, A.; Krahn, H.; Powell, M. *Contrasting fortunes across the life course: non-standard work among women and men in Canada and the United Kingdom*. Leicester: CLMS – Center for Labour Market Studies 1997 (Working paper, 17). Available from Internet: http://www.clms.le.ac.uk/publications/workingpapers/working_paper17.pdf [cited 31.8.2004].
- Firth, D.; Payne, C.; Payne, J. Efficacy of programmes for the unemployed: discrete time modelling of duration data from a matched comparison study. *Journal of the Royal Statistical Society*, 1999, Vol. 162, p. 111-120.
- Fitzenberger, B.; Prey, H. Assessing the impact of training on employment. *ifo-Studien*, 1997, Vol. 43, p. 71-116.
- Fitzenberger, B.; Prey, H. Beschäftigungs- und Verdienstwirkungen von Weiterbildungsmaßnahmen im ostdeutschen Transformationsprozeß: Eine Methodenkritik. In: Pfeiffer, F.; Pohlmeier, W. (eds) *Qualifikation, Weiterbildung und Arbeitsmarkterfolg*. Baden-Baden: Nomos-Verlag, 1998 (ZEW-Wirtschaftsanalysen, Vol. 31).
- Fitzenberger, B.; Prey, H. Berufliche Weiterbildung und die Stabilität der Beschäftigung. Eine Evaluation auf Basis des Sozioökonomischen Panels 1999. In: Beer, D. et al. (eds) *Die wirtschaftlichen Folgen der Aus- und Weiterbildung*. Munich: Hampp, 1999.
- Fitzenberger, B.; Prey, H. Evaluating public sector sponsored training in East Germany. *Oxford Economic Papers*, 2000, p. 497-520.
- Fontaine, C.; Monnier, E. Evaluation in France. In: Furubo, J.E.; Rist, R.C.; Sandahl, R. (eds) *International atlas of evaluation*. London: Transaction Publishers, 2002, p. 63-76.
- Foray, D.; Grübler, A. Technology and the environment: an overview. *Technological forecasting and social change*. 1997, Vol. 53, No 1, p. 3-13.
- Frenzel, H.J. Bildung und Partnerwahl. *ZUMA Nachrichten*, 1995, Vol. 19, No 36, p. 61-87.
- Fretwell, D.; Benus, J.; O'Leary, C.J. *Evaluating the impact of active labor market programs: results of cross-country studies in Europe and Central Asia*. Washington DC: The World Bank, 1999 (Social protection discussion paper, 9915).
- Friedlander, D.; Greenberg, D. H.; Robins, P.K. Evaluating government training programs for the economically disadvantaged. *Journal of Economic Literature*, 1997, Vol. 35, No 4, p. 1809-1855.
- Friedrich, W.; Schumacker, M. *Evaluierung der Umsetzung des Programmes Leonardo*

- da Vinci in Deutschland 1995-1999*. Study commissioned by BMBF. Kerpen: BMBF – Federal Ministry of Education and Research, January 2000.
- Furlong, A.; Hammer, T. *Youth unemployment and marginalisation in northern Europe*. Oslo: Nova, 2000.
- Furubo, J.E.; Rist, R.C.; Sandahl, R. (eds) *International atlas of evaluation*. London: Transaction Publishers, 2002.
- Furubo, J.E.; Sandahl, R. A diffusion perspective on global developments in evaluation. In: Furubo, J.E.; Rist, R.C.; Sandahl, R. (eds) *International atlas of evaluation* London: Transaction Publishers, 2002, p. 1-26.
- Gadamer, H.G. *Wahrheit und Methode. Grundzüge einer philosophischen Hermeneutik*. Third edition. Tübingen: JCB Mohr (Paul Siebeck), 1972.
- Ganzach, Y. Parents' education, cognitive ability, educational expectations and educational attainment: Interactive effects. *British Journal of Educational Psychology*, 2000, Vol. 70, p. 419-441.
- Garland, D. The culture of high crime societies. Some preconditions of recent 'law and order' policies. *British Journal of Criminology*. Oxford: University Press, 2000, Vol. 40, No 3, p. 347-375.
- Gash, V.; O'Connell, P.J. *The Irish graduate labour market: a six year follow-up survey of third level graduates from 1992*. Dublin: ESRI – Economic and Social Research Institute, 2000.
- Gemmel, N. Evaluating the impacts of human capital stocks and accumulation on economic growth: some new evidence. *Oxford Bulletin of Economics and Statistics*, 1996, Vol. 58, No 1, p. 9-28.
- Gerfin, M.; Lechner, M. *Microeconomic evaluation of the active labour market policy in Switzerland*. Mannheim: ZEW, 2000 (Discussion paper, 00-24).
- Giddens, A. *Positivism and sociology*. London: Heinemann Educational Books, 1975.
- Glaeser, E.L. et al. Measuring trust. *Quarterly Journal of Economics*, 2000, Vol. 115, Issue 3, p. 811-846.
- Glaser, B.G.; Strauss, A.L. *The discovery of grounded theory: strategies for qualitative research*. Aldine de Gruyter, Hawthorne, 1999.
- Glossary of evaluation terms*. Available from Internet: http://www.evaluation.org.uk/Pub_library/Glossary.htm [cited 17.9.2004].
- Goldthorpe, J.H. *Social mobility and class structure in modern Britain*. Oxford: Clarendon Press, 1980.
- Götz, K. *Zur Evaluierung beruflicher Weiterbildung. Band 1. Theoretische Grundlagen. Third edition*. Weinheim: Deutscher Studien Verlag, 1999.
- Goux, D.; Maurin, E. Education, expérience et salaire: tendances récentes et évolution de long terme. *Economie et prévision*, 1994, Vol. 116, p. 155-178.
- Graff, M. Zur Bedeutung der Bildung im Prozess der wirtschaftlichen Entwicklung [The significance of education within the process of economic development]. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 1996, Vol. 48, No 2, p. 274-295.
- Graham, S. Global grids of glass. *Urban Studies*, 1999, Vol. 36, No 5/6, p. 929-949.
- Gray, A.; Jenkins, B. Policy and program evaluation in the United-Kingdom: a reflective State? In: Furubo, J.E.; Rist, R.C.; Sandahl, R. (eds) *International atlas of evaluation*. London: Transaction Publishers, 2002, p. 129-153.
- Green, A.; Preston, J. Education and social cohesion: re-centering the debate. *Peabody Journal of Education*, 2001, Vol. 76, No 3-4, p. 247-284.
- Green, A.; Preston, J.; Malmberg, L.-E. Non-material benefits of education, training and skills at a macro level. In: Descy, P.; Tessaring, M. (eds) *Impact of education and training. Third report on vocational training research in Europe: background report*. Luxembourg: EUR-OP, 2004 (Cedefop Reference series, 54).
- Green, A.; Sakamoto, A. Models of high skills in national competition strategies. In: Brown, P.; Green, A.; Lauder, H. (eds) *High skills: globalisation, competitiveness and skill formation*. Oxford: University Press, 2001.

- Green, F.; Hoskins, M.; Montgomery, S. *The effects of training, further education and YTS on the earnings of young employees*. Leicester: Faculty of Social Science, 1994 (Discussion paper, 10).
- Gridley, G. et al. Is there a healthy worker effect for cancer incidence among women in Sweden? *American Journal Industrial Medicine*, 1999, Vol. 36, p. 93-99.
- Griliches, Z. Education, human capital and growth: a personal perspective. *Journal of Labor Economics*, 1997a, Vol. 15, No 1, Part 2, p. S330-S344.
- Griliches, Z. *The Simon Kuznets memorial lectures*. Draft, October 1997b.
- Groot, W. Productivity effects of enterprise-related training. *Applied Economic Letters*. New York: Routledge, 1999, Vol. 6, No 6, p. 369-371.
- Grootings, P. VET in transition: an overview of changes in three East European countries. *European Journal of Education*, 1993, Vol. 28, No 2.
- Grossman, M.; Kaestner, R. Effects on education on health. In: Behrman, J.R.; Stacey, N. (eds) *The social benefits of education*. Ann Arbor: University of Michigan Press, 1997.
- Grubb, W.N. *Evaluating job training programmes in the United States: evidence and explanations*. Berkeley: NCRVE – National Center for Research in Vocational Education, 1995 (Technical assistance report, MDS-1047).
- Grubb, W.N.; Ryan, P. *The roles of evaluation for vocational education and training: plain talk in the field of dreams*. Geneva: ILO – International Labour Organisation, 1999.
- Guba, E.G.; Lincoln, Y.S. *Fourth generation evaluation*. London: Sage Publications, 1989.
- Guba, E.G.; Lincoln, Y.S. Paradigmatic controversies, contradictions and emerging confluences. In: Denzin, N.K; Lincoln, Y.S. (eds) *Handbook of qualitative research. Second edition*. London: Sage Publications, 2000.
- Guiso, L.; Sapienza, P.; Zingales, L. *The role of social capital in financial development*. Cambridge: National Bureau of Economic Research, 2000 (NBER Working paper, 7563).
- Gunnarsson, G.; Mellander, E.; Savvidou, E. *Is human capital the key to the IT productivity paradox?* Stockholm: IUI – Research Institute of Industrial Economics, 2001 (Working paper, 551).
- Hægeland, T.; Klette, T.J.; Salvanes, K.G. Declining returns to education in Norway? Comparing estimates across cohorts, sectors and over time. *Scandinavian Journal of Economics*, 1999, Vol. 101, No 4, p. 555-576.
- Hall, E.M.; Johnson, J.V.; Tsou, T.-S. Women, occupation, and risk of cardiovascular morbidity and mortality. *Occupational Medicine: State of the Art Reviews*, 1993, Vol. 8, p. 709-719.
- Hall, J.; Jones, C. Why do some countries produce so much more output per workers than others? *Quarterly Journal of Economics*, 1999, 114, 1, p. 83-116.
- Halman, L. Variations in tolerance levels in Europe: evidence from the eurobarometers and European values study. *European Journal on Criminal Policy and Research*. Dordrecht: Kluwer, 1994, 2-3, p. 15-38.
- Hammarström, M. *Varför inte högskola?: En longitudinell studie av olika faktorerens betydelse för studiebegåvade ungdomars utbildningskarriär*. Göteborgs: Institutionen för pedagogik, 1996 (Göteborg studies in educational sciences, 107).
- Hammond, C. *Learning to be healthy*. London: Centre for Research on the Wider Benefits of Learning, 2002 (Research brief, RCB07).
- Handley, D. et al. *Study on the opening of the Leonardo da Vinci programme to the countries of Central and Eastern Europe*. NCVQ: London, 1996.
- Hannan, D.F. *Education, vocational training and labour market transition (VTLMT) amongst lower level school leavers in 4 European countries*. Dublin: ESRI, 2001 (ESRI report).
- Hannan, D.F.; Werquin, P. Education and labour market change: the dynamics of education to work transitions in Europe. A review of the TSER programme. In: Descy, P.; Tessaring, M. (eds) *Training in Europe: second report on vocational training*

- research in Europe 2000: background report. Volume 3* Luxembourg: EUR-OP, 2001 (Cedefop Reference series, 3008).
- Hansson, B.M. *Marketable human capital investments: an empirical study of employer sponsored training*. Stockholm: School of Business, 2001 (Working paper).
- Hansson, B.; Johanson, U.; Leitner, K.H. The impact of human capital and human capital investments on company performance: evidence from literature and European survey results. In: Descy, P.; Tessaring, M. (eds) *Impact of education and training. Third report on vocational training research in Europe: background report*. Luxembourg: EUR-OP, 2004 (Cedefop Reference series, 54).
- Hanushek, E.A.; Kim, D. *Schooling, labor force quality and economic growth*. Cambridge: National Bureau of Economic Research, 1995 (NBER Working paper, 5399).
- Hanushek, E.A.; Kimko, D. Schooling, labor-force quality and the growth of nations. *American Economic Review*, 2000, Vol. 90, No 5, p. 1184-1208.
- Hara, V. *Evaluation of education for the information society*. Publication of Finnish Higher Education Evaluation Council. Helsinki: Edita, 2000.
- Harmon, C.; Walker, I. *The returns to education: a review of the evidence, issues and deficiencies in the literature*. Norwich: HSMO, 2001 (DfEE Research report, 254).
- Harmon, C.; Walker, I.; Westergaard-Nielsen, N. Returns to education in Europe. In: PURE. *Public funding and private returns to education*, 2001, p. 13-21 (Project final report series – TSER Area 2 – SOE2-CT98-2044).
- Harmon, M. et al. *Performance assessment in IEA's third international mathematics and science study (TIMSS)*. Chestnut Hill, MA: TIMSS International Study Centre, 1997. Available from Internet: <http://timss.bc.edu/timss1995i/PAreport.html> [cited 1.9.2004].
- Harms, J. Wirtschaftlichkeit unter Bedingungen des New Public Management – unter besonderer Berücksichtigung des Schulwesens. In: Weiß, M.; Weishaupt, H. (eds) *Bildungsökonomie und neue Steuerung*. Frankfurt/M.: Peter Lang, 2000, p. 133-148.
- Hatcher, R. Class differentiation in education: rational choices? *British Journal of Sociology of Education*. Philadelphia: Carfax publishing, 1998, Vol. 19, No 1, p. 5-24.
- Haughton, G. et al. *Prototype employment zones: a qualitative and contextual evaluation*. Sheffield: DfEE – Department for Education and Employment, 2000 (Research report, 232).
- Healy, T. Investing in human capital – the OECD view. In: Weiß, M.; Weishaupt, H. (eds) *Bildungsökonomie und neue Steuerung*. Frankfurt/M.: Peter Lang, 2000, p. 19-29.
- Heinrich, G.; Hildebrandt, V. *Public and private returns to education in the European Union – an appraisal*. Luxembourg: European Investment Bank, 2001.
- Heinz, W.R. *From education to work: cross-national perspectives*. Revision of papers delivered at a conference at the University of Toronto, April 1996. Cambridge: University Press, 1999.
- Heise, M.; Meyer, W. The benefits of education, training and skills from an individual life-course perspective with a particular focus on life-course and biographical research. In: Descy, P.; Tessaring, M. (eds) *Impact of education and training. Third report on vocational training research in Europe: background report*. Luxembourg: EUR-OP, 2004 (Cedefop Reference series, 54).
- Helliwell, J.F. *Do borders matter for social capital? Economic growth and civic culture in U.S. states and Canadian provinces*. Cambridge: National Bureau of Economic Research, 1996a (NBER Working paper, 5863).
- Helliwell, J.F. *Economic growth and social capital in Asia*. Cambridge: National Bureau of Economic Research, 1996b (NBER Working paper, 5470).
- Helliwell, J.F.; Putnam, R.D. *Education and social capital*. Cambridge: National Bureau of Economic Research, 1999 (NBER Working paper, 7121).

- Hellwig, W. et al. Evaluation of EU and international programmes and initiatives promoting mobility: selected case studies. In: Descy, P.; Tessaring, M. (eds) *Evaluation of systems and programmes. Third report on vocational training research in Europe: background report*. Luxembourg: EUR-OP, 2004 (Cedefop Reference series, 57).
- Hillmert, S. *Ausbildungssysteme und Arbeitsmarkt. Lebensverläufe in Deutschland and Großbritannien im Kohortenvergleich*. Wiesbaden: Westdeutscher Verlag, 2001.
- Hillmert, S. Stabilität und Wandel des 'deutschen Modells': Lebensverläufe im Übergang zwischen Schule und Beruf. In: Wingens, M.; Sackmann, R. (eds) *Bildung und Beruf. Ausbildung und struktureller Wandel in der Wissensgesellschaft*. München: Juventa, 2002.
- Holford, J.; van der Veen, R. *Lifelong learning governance and active citizenship in Europe*. Guildford: University of Surrey, 2003. Available from Internet: <http://www.surrey.ac.uk/Education/ETGACE/Final-Report-Screen-version.pdf> [cited 1.9.2004].
- Hood, C. A public management for all seasons. *Public Administration*, 1991, Vol. 69, No 1, p. 3-19.
- House, E.R. Assumptions underlying evaluation models. In: Madaus, G.F.; Scriven, M.; Stufflebeam, D.L. *Evaluation models viewpoints on educational and human services evaluation*. The Hague: Kluwer-Nijhoff Publishing, 1983.
- Hradil, S.; Immerfall, S. (eds) *Die westeuropäischen Gesellschaften im Vergleich*. Opladen: Leske + Budrich, 1997.
- Hübler, O. Berufliche Weiterbildung und Umschulung in Ostdeutschland – Erfahrungen und Perspektiven. In: Pfeiffer, F.; Pohlmeier, W. (eds) *Qualifikation, Weiterbildung und Arbeitsmarkterfolg*. Baden-Baden: Nomos-Verlag, 1998 (ZEW-Wirtschaftsanalysen, Vol. 31).
- Huggins, R.; Izushi, H. *World knowledge competitiveness index 2002: benchmarking the globe's high performing regions*. Cardiff: Robert Huggins Business and Economic Policy Press, 2002.
- Huinink, J. Bildung und Familienentwicklung im Lebenslauf. *Zeitschrift für Erziehungswissenschaft*, 2000, Vol. 3, No 2, p. 209-227.
- Hujer, R.; Caliendo, M.; Radić, D. Methods and limitations of evaluation and impact research. In: Descy, P.; Tessaring, M. (eds) *The foundations of evaluation and impact research. Third report on vocational training research in Europe: background report*. Luxembourg: EUR-OP, 2004a (Cedefop Reference series, 58).
- Hujer, R.; Caliendo, M.; Zeiss, C. Macroeconometric evaluation of active labour-market policy – a case study for Germany. In: Descy, P.; Tessaring, M. (eds) *Impact of education and training. Third report on vocational training research in Europe: background report*. Luxembourg: EUR-OP, 2004b (Cedefop Reference series, 54).
- Hujer, R. et al. Kurz- und langfristige Effekte von Weiterbildungsmaßnahmen auf die Arbeitslosigkeitsdauer in Westdeutschland. In: Pfeiffer, F.; Pohlmeier, W. (eds) *Qualifikation, Weiterbildung und Arbeitsmarkterfolg*. Baden Baden: Nomos-Verlag, 1998, p. 197-221 (ZEW-Wirtschaftsanalysen, Vol. 31).
- Hujer, R. et al. Analyzing the effects of on-the-job vs. off-the-job training. In: Hujer, R.; Maurer, K.O.; Wellner, M. Estimating the effect of vocational training on unemployment duration in West Germany: a discrete hazard rate model with instrumental variables. *Jahrbücher für Nationalökonomie und Statistik*, 1999a, Vol. 218/5+6, p. 619-646.
- Hujer, R. Maurer, K.O.; Wellner, M. The effects of public sector sponsored training on unemployment duration in West Germany: a discrete hazard rate model based on a matched sample. *ifo-Studien*, 1999b, Vol. 3, p. 371-410.
- Hujer, R; et al. Analyzing the effects of on-the-job vs. off-the-job training on unemployment duration in West Germany. In: Bellmann, L; Stainer, V. (eds) *Panelanalysen zu Lohnstruktur, Qualifikation und Beschäftigungsdynamik*. Nuremberg: IAB, 1999c, p. 203-237 (Beiträge zur Arbeitsmarkt- und Berufsforschung Serie, 229).

- Hujer, R.; Wellner, M. *The effects of public sector sponsored training on individual employment performance in East Germany*. Bonn: IZA, 2000a (Discussion paper, 141).
- Hujer, R.; Wellner, M. *Berufliche Weiterbildung und individuelle Arbeitslosigkeitsdauer in West- und Ostdeutschland: eine mikroökonomische Analyse. Mitteilungen für Arbeitsmarkt- und Berufsforschung*, 2000b, No 3, p. 405-417.
- Hullen, G. *Lebensverläufe in West- und Ostdeutschland. Längsschnittanalysen des deutschen Family and Fertility Surveys*. Leverkusen: Leske und Budrich, 1998.
- Hullen, G. *The effects of education and employment on marriage and first birth*. Paper in the framework of the FFS Flagship conference 'Partnership and fertility – a revolution', Brussels, 29-31 May 2000.
- Human Resources Development Canada. *Lessons learned: effectiveness of employment-related programmes for youth*. Ottawa: HRDC – Human Resources Development Canada, Ottawa, June 1997.
- Ichniowski, C.; Shaw, K.; Prenzushi, G. *The effects of human resource management practices on productivity*. Cambridge: National Bureau of Economic Research, 1995 (NBER Working paper, W5333).
- Inglehart, R. et al. (eds) *Human beliefs and values. A cross-cultural sourcebook based on the 1999-2002 values surveys*. Mexico City: Siglo XXI, 2004.
- IPF. *The human capital survey 2002*. Uppsala: IPF – Institute of Personnel and Corporate Development, 2002.
- Islam, N. Growth empirics: a panel data approach. *Quarterly Journal of Economics*, 1995, Vol. 110, No 4, p. 1127-1170.
- IZA. Evaluation of labor market policy programs – Germany and Europe must catch up with the US. *IZA Compact*, December 2000, p. 1-3.
- IZA. Evaluation is essential for a reform of active labor market policy. *IZA Compact*, October 2001, p. 1-4.
- Izushi, H.; Huggins, R. Empirical analysis of human capital development and economic growth in European regions. In: Descy, P.; Tessaring, M. (eds) *Impact of education and training. Third report on vocational training research in Europe: background report*. Luxembourg: EUR-OP, 2004 (Cedefop Reference series, 54).
- Jahnukainen, M. *Social exclusion and dropping out of education*. In: Visser, J.; Daniels, H.; Cole, T. (eds) *Emotional and behavioural difficulties in mainstream schools*. Amsterdam: Elsevier, 2001, Vol. 1, p. 1-12.
- Jann, W. Politikfeldanalyse. In: Nohlen, D. *Lexikon der Politik. Bd. 2, Politikwissenschaftliche Methoden*. München: Beck, 1994.
- Jarvis, V.; O'Mahony, M.; Wessels, H. *Product quality, productivity and competitiveness: a study of the British and German ceramic tableware industries*. London: National Institute of Economic and Social Research, 2002 (Occasional paper, 55).
- Jenkins, H. *Education and production in the United Kingdom*. Oxford: Nuffield College, 1995 (Economic discussion paper, 101).
- Jensen, P. Nielsen, M. S.; Rosholm, M. *The effects of benefit, incentives, and sanctions on youth unemployment*. Århus: Center of Labour Market and Social Research – CLS, 1999 (Working paper, 99-05).
- Jenson, J. *Mapping social cohesion: the state of Canadian research*. Ottawa: Renouf Publishing, 1998 (CPRN Study, No F/03). Available from Internet: <http://www.cprn.com/en/doc.cfm?doc=180> [cited 6.9.2004].
- Jorgenson, D.W.; Fraumeni, B. Investment in education and US economic growth. *Scandinavian Journal of Economics*, 1992, Vol. 94, p. S51-S70.
- Judson, R. Economic growth and investment in education: how allocation matters. *Journal of Economic Growth*, 1998, Vol. 3, No 4, p. 337-360.
- Kaiser, L.; Siedler, T. *Exits from unemployment spells in Germany and the United Kingdom*. Colchester: EPAG – European Panel Analysis Group, 2000 (Working paper, 7).
- Katz, E.; Ziderman, A. Investment in general training: the role of information and labour mobility. *Economic Journal*, 1990, Vol. 100, p. 1147-1158.

- Kazamaki Ottersten, E.; Lindhl, T.; Mellander, E. *Cost and productivity effects of firm financed training*. Uppsala: Industrial Institute for Economic and Social Research, 1996 (Working paper, 455).
- Kazamaki Ottersten, E.; Lindhl, T.; Mellander, E. Evaluating firm training, effects on performance and labour demand. *Applied Economic Letters*, 1999, Vol. 6, No 7, p. 431-427.
- Kelleher, M. *Vet as a learning organisation*. Usk: CRED, 2001 (Working paper, Forum Network).
- Kelly, M. Inequality and crime. *The Review of Economics and Statistics*. Cambridge, MIT Press, 2000, Vol. 82, No 4, p. 530-539.
- Keogh, H.; Downes, T. *VTO spells success*. Dublin: Department of Education and Science, 1998.
- Kettunen, J. The effects of education on the duration of unemployment. *Review of Labour Economics and Industrial Relations*, 1994, Vol. 8, No 2, p. 331-352.
- Kieselbach, T. (ed.) *Youth unemployment and social exclusion: a comparison of six European countries*. Bremen: University of Bremen, 2000 (Psychology of social inequality series, Vol. 9, Yuseder Publication, No 1).
- Klenow, P.; Rodriguez-Clare, A. The neo-classical revival in growth economics: has it gone too far? In: Bernanke, B.S.; Rotemberg, J.J. (eds) *NBER Macroeconomics Annual 1997*. Cambridge: MIT Press, 1997, p. 73-102.
- Klijzing, E. *Globalization and the changing role of education in the process of entry into the labour market: a cross-national comparison of five countries*. Bielefeld: Faculty of Sociology, 2000 (Globalife Working paper, 12).
- Kluge, F. *Etymologisches Wörterbuch der deutschen Sprache* (bearb. von Elmar Seebold). 23rd edition. Berlin: Walter de Gruyter, 1999.
- Knack, S.; Keefer, P. Does social capital have an economic payoff? A cross-country investigation. *Quarterly Journal of Economics*, 1997, Vol. 112, No 4, p. 1251-1288.
- Koivuluhta, M. *Ammatti – intressit ja ura: pohjoiskarjalaisen ammattikoulutetun ikäluokan seuranta vuosina 1975-1991*. Joensuu: Joensuun yliopisto, 1999.
- Konietzka, D. Berufliche Aus- und Fortbildung in der Lebensverlaufsperspektive. Ein Vergleich des Ausbildungsverhaltens sechs westdeutscher Geburtskohorten. *Zeitschrift für Pädagogik*, 1999, Vol. 45, No 6, p. 807-831.
- Konietzka, D. Die soziale Differenzierung der Übergangsmuster in den Beruf. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 2002, Vol. 45, No 4, p. 645-673.
- Kortteinen, M.; Tuomikoski, H. Miten työttömät selviytyvät? [How do the unemployed survive?], *Yhteiskuntapolitiikka*, 1998, Vol. 63/1, p. 5-13.
- Kraus, F.; Puhani, P.A.; Steiner, V. *Do public works programmes work? Some unpleasant results from the East German experience*. Mannheim: ZEW, 1998 (Discussion paper, 98-07).
- Krekel, E.M.; Bardeleben, R.V.; Beicht, U. Bildungscontrolling: Hintergrund, Bedeutung und Definition. In: Krekel, E.M. et al. (eds) *Controlling in der beruflichen Weiterbildung im europäischen Vergleich. Berichte zur beruflichen Bildung [A European Comparison of Controlling in Corporate Continuing Training]*. Bielefeld: Bertelsmann, 2001, Vol. 20.
- Kromrey, H. Evaluation – ein vielschichtiges Konzept. *Sozialwissenschaften und Berufspraxis*, 2001, Vol. 2, p. 105-132.
- Krueger, A.B.; Lindahl, M. *Education for growth: why and for whom?* Princeton: Princeton University, 1998 (Working paper, 429).
- Krueger, A.B.; Lindahl, M. Education for growth: why and for whom? *Journal of Economic Literature*, 2001, Vol. 39, No 4, p. 1101-1136.
- Krumm, V.; Weiß, S. Ungerechte Lehrer. *Salzburger Beiträge zur Erziehungswissenschaft*, 2000, Vol. 1, No 4, p. 60-79.
- Kvale, S. *Interview*. Gylling: Hans Reitzels Forlag, 2003.
- Kyriacou, G. *Level and growth effects of human capital: a cross-country study of the convergence hypothesis*. New York: CV Starr Center, 1991 (Working paper, 91-26).

- La Porta, R. et al. Trust in large organisations. *American Economic Review*, 1997, Vol. 87, No 2, p. 333-338.
- Lalive, R.; Van Ours, J. C.; Zweimüller, J. *The impact of active labor market programmes and benefit entitlement rules on the duration of unemployment*. Bonn: IZA, 2000 (Discussion paper, 149).
- Landau, D. Government expenditure and economic growth: a cross-country study. *Southern Economic Journal*, January 1983, p. 783-792.
- Lange, E. Zur Entwicklung und Methodik der Evaluationsforschung in der Bundesrepublik Deutschland. *Zeitschrift für Soziologie*, 1983, Vol. 12, No 3, p. 253-270.
- Larsson, L. *Evaluation of Swedish youth labour market programmes*. Uppsala: IFAU – Institute for labour market policy evaluation, 2000 (Working paper series, 2000:1).
- Lasheras, C. et al. Effects of education on the quality of life, diet, and cardiovascular risk factors in an elderly Spanish community population. *Experimental Aging Research*, 2001, Vol. 27, p. 257-270.
- Lauer, C. *Family background, cohort and education. A French-German comparison*. Mannheim: Centre for European Economic Research, 2002 (ZEW Discussion paper, 02-12).
- Laursen, K.; Foss, N.J. New human resource management practices, complementarities and the impact on innovation performance. *Cambridge Journal of Economics*, 2003, Vol. 27, No 2, p. 243-263.
- Lave, J.; Wenger, E. *Situated learning. Legitimate peripheral participation*. Cambridge: University Press, 1991.
- Le Grand, J. *The quasi-market experiments in public service delivery: did they work?* Paper for presentation at Pontignano conference 6-8 April 2001. London: London School of Economics, 2001. Available from internet: <http://www.econ-pol.unisi.it/welfare/legrand.pdf> [cited 6.9.2004].
- Le Grand, J.; Bartlett, W. (eds) *Quasi-markets and social policy*. London: Macmillan Press, 1993.
- Lechner, M. Earnings and employment effects of continuous off-the-job training in East Germany after unification. *Journal of Business Economic Statistics*, 1999a, Vol. 17, p. 74-90.
- Lechner, M. The effects of enterprise-related continuous vocational training in East Germany on individual employment and earnings. *Annals d'Économie et de Statistique*, 1999b, Vol. 55-56, p. 97-128.
- Lechner, M. An evaluation of public sector sponsored continuous vocational training programmes in East Germany. *Journal of Human Resources*, 2000, Vol. 35, No 2, p. 347-375.
- Lee, M.R. Concentrated poverty, race and homicide. *Sociological Quarterly*, 2000, Vol. 41, Issue 2, p. 189-206.
- Lee, J.W.; Barro, R.J. Schooling quality in a cross-section of countries. *Economica*, 2001, Vol. 68, Issue 272, p. 465-488.
- Lee, J.W.; Lee, T.H. Human capital and economic growth. Tests based on the international evaluation of educational achievement. *Economics Letters*, 1995, 47, p. 219-225.
- Leeuw, F.L. Evaluation in Europe. In: Stockmann, R. (ed.) *Evaluationsforschung: Grundlagen und ausgewählte Forschungsfelder*. Opladen: Leske und Budrich, 2000, p. 57-76 (Sozialwissenschaftliche Evaluationsforschung, 01).
- Leeuw, F.L. Evaluation in Europe 2000: challenges to a growth industry. *Evaluation*, 2002, Vol. 8, Issue 1, p. 5-12.
- Leiponen, A. *Competence, innovation and profitability of firms*. Helsinki: Research Institute of the Finnish Economy, 1996a (ETLA Working paper, 563).
- Leitner, K.H. *Intangible resources and firms performance: empirical evidence from Austrian SMEs*. Paper prepared for the 16th Nordic Academy of Management Meeting, Uppsala 16-18 August 2001.
- Lemkow, L. et al. Youth unemployment and social exclusion in Spain. In: Kieselbach, T. (ed.) *Youth unemployment and social exclusion: a comparison of six European countries*. Bremen: University of Bremen, 2000 (Psychology of social inequality

- series, Vol. 9, Yuseder Publication, No 1).
- Lerman, R.I. *Employment and training programmes for out-of-school youth*. Washington DC: Urban Institute, 1997 (mimeo).
- Levine, R.; Renelt, D. A sensitivity analysis of cross-country growth regressions. *American Economic Review*, 1992, Vol. 82, No 4, p. 942-963.
- Liefbroer, A. *Transition from youth to adulthood in the Netherlands*. Bamberg: Department of Sociology, 2002 (Globalife working paper, 22).
- Lin, N. *Social capital theory: a theory of social structure and social action*. Cambridge: University Press, 2001.
- Lockwood, D. *Solidarity and Schisms, 'the problem of disorder' in Durkheimian and Marxist sociology*. Oxford: Clarendon Press, 1992.
- Loewenstein, M.A.; Spletzer, J.R. Dividing the cost and returns to general training. *Journal of Labor Economics*, 1998, Vol. 16, No 1, p. 142-171.
- Loewenstein, M.A.; Spletzer, J.R. General and specific training: evidence and implications. *Journal of Human Resources*, 1999, Vol. 34, No 4, p. 710-733.
- Lucas, R.E. On the mechanics of economic development. *Journal of Monetary Economics*, 1988, Vol. 22, No 1, p. 3-42.
- Lundvall, B.Å. *Europe and the learning economy: on the need for reintegrating the strategies of firms, social partners and policy makers*. Ålborg: Department of Business Studies, 2000.
- Lundvall, B.Å.; Borrás, S. *The globalising learning economy: implications for innovation policy*. Report based on contributions from seven projects under the TSER programme. DGXII, European Commission, December 1997.
- Luschei, F.; Trube, A. Evaluation und Qualitätsmanagement in der Arbeitsmarktpolitik – Einige systematische Vorüberlegungen und praktische Ansätze zur lokalen Umsetzung. *MittAB*, 2000, No 3, p. 533-549.
- Lynch, L.; Black, S. *Beyond the incidence of training: evidence from a national employers survey*. Cambridge: National Bureau of Economic Research, 1995 (NBER working paper, 5231).
- MacDuffie. Human resources bundles and manufacturing performance: organisational logic and flexible production systems in the world auto industry. *Industrial and Labor Relations Review*, 1995, Vol. 48, No 2, p. 197-221.
- Mackenbach, J.P. et al. *Socioeconomic determinants of healthy aging. From description to explanation (SEdHA)*. Rotterdam: Erasmus University, Department of Public Health, 1999 (Unpublished proposal submission form).
- Maglen, L.; Hopkins, S. *Linking VET to productivity differences: an evaluation of the Prais program and its implications for Australia*. Victoria: Monach University, 1999 (CEET Working paper, 18).
- Magoula, T.; Psacharopoulos, G. Schooling and monetary rewards in Greece: an over-education false alarm. *Applied Economics*, 1999, Vol. 31, Issue 12, p. 1589-1597.
- Mankiw, N.G.; Romer, D.; Weil, D.N. A contribution to the empirics of economic growth. *Quarterly Journal of Economics*, 1992, Vol. 107, No 2, p. 407-437.
- Marmot, M; Wilkinson, J. (eds) *The social determinants of health*. Oxford: University Press, 1999.
- Martin, J.P. What works among active labour market policies: evidence from OECD countries' experiences, In: Debelle, G.; Borland, J. (eds) *Unemployment and the Australian labour market*. Proceedings of a conference, 1998. Available from Internet: <http://ideas.repec.org/p/rba/rbaacv/1998-17.html> [cited 2.9.2004].
- Martin, J.P.; Grubb, D. *What works and for whom: a review of OECD countries' experiences with active labour market policies*. Uppsala: IFAU – Institute for labour market policy evaluation, 2001 (Working paper series, 2001:14).
- Mason, G.; Keltner, B.; Wagner, K. *Productivity, technology and skills in banking: commercial lending in Britain, the United States and Germany*. London: National Institute of Economic and Social Research, 1999 (Discussion paper, 159).

- Maxwell, J. *Social dimension of economic growth*. Alberta: University of Alberta, 1996 (Eric John Hanson memorial lecture series, Vol. VIII).
- Mayer, K.U. Lebenslaufforschung. In: Voges, W. (ed.) *Methoden der Biographie- und Lebenslaufforschung*. Opladen: Leske und Budrich, 1987.
- Mayer, K.U. Ausbildungswege und Berufskarrieren. In: Bundesinstitut für Berufsbildung (ed.) *Forschung im Dienst von Praxis und Politik*. Bielefeld: Bertelsmann, 1996.
- Mayer, K.U.; Brückner, H. *Lebensverläufe und gesellschaftlicher Wandel. Konzeption, Design und Methodik der Erhebung von Lebensverläufen der Geburtsjahrgänge 1954-1956 und 1959-1961 [Life Course and Social Change. Conceptualization, Design, and Methods for the Collection of Life Course Data of the Birth Cohorts 1954-56 and 1959-61]*. Berlin: Max-Planck-Institut für Bildungsforschung, 1995 (Materialien aus der Bildungsforschung No 48).
- McCracken, M. *Social cohesion and macroeconomic performance*. Centre for the Study of Living Standards (CSLS). Conference on the state of living standards and the quality of life, held on 30-31 October 1998 in Ottawa, Ontario.
- McIntosh, S.; Steedman, H. *Low skills: a problem for Europe*. Luxembourg: EUR-OP, 1999 (Project final report series – TSER Area 2).
- McMahon, W. Conceptual framework for the analysis of the social benefits of lifelong learning. *Education Economics*, 1998, Vol. 6, No 3, p. 309-346.
- McMahon, W. *Education and development: measuring the social benefits*. Oxford: University Press, 2000.
- Melck, A. From normative theory to political economy: public choice in the educational system. *Zeitschrift für Wirtschaftspolitik*, 1991, Vol. 40, No 3, p. 271-283.
- Meritum. *Guidelines for managing and reporting on intangibles*. Cañibano, L. et al. (eds) Fundación Airtel Móvil, 2002.
- Merrill, B. Biographies as collective experience? In: *ESREA: The Biography and Life History Network Conference, submitted paper*; Geneva, 7-9 March 2002.
- Mertens, D.M. *Research methods in education and psychology: integrating diversity with quantitative and qualitative approaches*. Thousands Oaks, CA: Sage, 1998.
- Mertens, D.M. Institutionalizing evaluation in the United States of America. In: Stockmann, R. (ed.) *Evaluationsforschung. Grundlagen und ausgewählte Forschungsfelder*. Opladen: Leske and Budrich, 2000, p. 41-56 (schafftliche Evaluationsforschung No 01).
- Meulemann, H. Schullaufbahnen, Ausbildungskarrieren und die Folgen im Lebensverlauf. Der Beitrag der Lebenslaufsforschung zur Bildungssoziologie. In: Mayer, K.U. (ed.) *Lebensverläufe und sozialer Wandel*. Opladen: Westdeutscher Verlag, 1990.
- Michie, J.; Sheehan, M. HRM practices, R&D expenditure and innovative investment: evidence from the UK's 1990 workplace industrial relations survey. *Industrial and Corporate Change*. Oxford: University Press, 1999, Vol. 8, Issue 2, p. 211-234.
- Mincer, J. *Schooling, experience and earnings*. New-York: National Bureau of Economic Research, 1974.
- Mincer, J. Education and unemployment. In: Mincer, J. (ed.) *Studies in human capital*. Cambridge: Edward Elgar Publishing, 1993, p. 212-238.
- Mingat, A.; Tan, J.P. *The full social returns to education: estimates based on countries' economic growth performance*. Human Capital Development Papers. Washington: The World Bank, 1996.
- Modernising Government*. Presented to Parliament by the Prime Minister and the Minister for the Cabinet Office by Command of Her Majesty. London: HMSO, 1999. Available from Internet: <http://www.archive.official-documents.co.uk/document/cm43/4310/4310.htm> [cited 8.9.2004].
- Molsosa, J. Evaluation activities in the European Commission. In: Descy, P.; Tessaring, M. (eds) *Evaluation of systems and programmes. Third report on vocational training research in Europe*:

- background report*. Luxembourg: EUR-OP, 2004 (Cedefop Reference series, 57).
- Morrow, R.A.; Torres, C.A. *Social theory and education: a critique of theories of social and cultural reproduction*. Albany: State University of New York Press, 1995.
- Mortensen, N. Mapping system integration and social integration. In: Gough, I.; Olofsson, G. (eds) *Capitalism and social cohesion: essays on exclusion and integration*. Basingstoke: Macmillan, 1999.
- Müller, W.; Gangl, M.; Scherer, S. Übergangsstrukturen zwischen Bildung und Beschäftigung. In: Wingens, M.; Sackmann, R. (eds) *Bildung und Beruf*. Weinheim: Deutscher Universitätsverlag, 2002.
- Müller, W.; Shavit, Y. Bildung und Beruf im institutionellen Kontext. Eine vergleichende Studie in 13 Ländern. *Zeitschrift für Erziehungswissenschaft*, 1998, No 1, p. 501-533.
- Murphy, K.M.; Shleifer, A.; Vishny, R.W. The allocation of talent: implications for Growth. *Quarterly Journal of Economics*, 1991, Vol. 106, No 2, p. 503-530.
- NA/NTF – LdV National Agency/National Training Fund. *Evaluation of the first phase of the LdV programme in the Czech Republic (1996-1999) and valorisation of its results*. Prague: NA/NTF, 2002.
- NCU/NTF – National Coordination Unit/National Training Fund. *Report on the LdV programme: Czech Republic*. Prague: NCU/NTF, 1999.
- Nehru, V.; Dhareshwar, A. A new database on physical capital stock: sources, methodology and results. *Revista de Analisis Economico*, 1993, Vol. 8, No 1, p. 37-59.
- Nelson, R.R.; Phelps, E.S. Investment in humans, technological diffusion and economic growth. *American Economic Review*, 1966, Vol. 56, No 2, p. 69-75.
- Nicaise, I. Giving fish or teaching to fish? A cost-benefit analysis of Belgian employment-training projects for minimum income recipients. *Public Finance and Management*, 2000, Vol. 2, No 2.
- Nicaise I., Bollens J. Training and employment opportunities for disadvantaged persons. In: Tessaring, M. (ed.) *Vocational education and training – the european research field: background report, vol. 2*. Luxembourg: EUR-OP, 1998, p. 121-153 (Cedefop Reference Document, 3002).
- Nieuwenhuis, L.; Shapiro, H. Evaluating systems reform in vocational education and training: learning from Danish and Dutch cases. In: Descy, P.; Tessaring, M. (eds) *Evaluation of systems and programmes. Third report on vocational training research in Europe: background report*. Luxembourg: EUR-OP, 2004 (Cedefop Reference series, 57).
- Noguera, C.; Golsch, K.; Bonmati, A. *Globalization and occupational mobility in adult male job careers in Spain*. Bamberg: Department of Sociology, 2002 (Globalife working paper, 35).
- Nonaka, I.; Takeuchi, H. *The knowledge creating company: how Japanese companies create the dynamic of innovation*. New York: Oxford University Press, 1995.
- Nummenmaa, A.R. *Koulutus, sukupuoli ja elämäntilanne. Nuoruudesta aikuisuuteen yhteiskunnallisessa muutoksessa*. Helsinki: Työministeriö, 1996 (Työpoliittisia tutkimuksia, 149).
- NUTEK. *Företag i förändring: Lärandestrategier för ökad konkurrenskraft [Enterprises in transformation: learning strategies for improved competitive power]*. Stockholm: Närings och Teknikutvecklingsverket, 2000.
- O'Connell, P.; McGinnity, F. What works, who works? The employment and earnings effects of active labour market programmes among young people in Ireland. *Work, Employment and Society*, 1997, Vol. 11, No 4, p. 639-661.
- OECD. *Employment Outlook: 1993*. Paris: OECD, 1993.
- OECD. *Enhancing the effectiveness of active labour market policies*. The OECD jobs strategy. Paris: OECD, 1996.
- OECD. *Labour market policies: new challenges. Enhancing the effectiveness of active labour market policies: a streamlined employment service*. Meeting of the Employment, Labour and Social Affairs Committee at Ministerial level held at the

- Château de la Muette in Paris on 14 and 15 October 1997. Paris: OECD, 1997.
- OECD. *Human capital investment – an international comparison*. Paris: OECD, 1998.
- OECD. *Transition from initial education to working life – making transition work*. Paris: OECD, 2000.
- OECD. *Knowledge and skills for life. First result from PISA 2000*. Paris: OECD, 2001a.
- OECD. *Employment outlook: 2001*. Paris: OECD, 2001b.
- OECD. *Starting strong: early childhood education and care*. Paris: OECD, 2001c.
- OECD. *The well-being of nations: the role of human and social capital*. Paris: OECD, 2001d.
- OECD. *Beyond rhetoric: adult learning policies and practices*. Draft Highlights. Paris: OECD, 2002.
- OECD. *Beyond rhetoric: adult learning policies and practices*. Paris: OECD, 2003a.
- OECD. *Education at a glance*. Paris: OECD, 2003b.
- OECD; Statistics Canada. *Littératie et société du savoir: nouveaux résultats de l'enquête internationale sur les capacités de lecture et d'écriture des adultes*. Paris: OCDE, 1997.
- OECD; Statistics Canada. *Literacy in the information age: final report of the international adult literacy survey (IALS)*. Paris: OCDE, 2000.
- Osberg, L.; Sharpe, A. *Comparisons of trends in GDP and economic well-being: the impact of social capital*. Paper presented at the international symposium on the contribution of human and social capital to sustained economic growth and well-being, organised by the OECD and HRDC, Quebec City, Canada, 19-21 March 2000.
- Pannenberg, M. *Weiterbildungsaktivitäten und Erwerbsbiographie – Eine empirische Analyse für Deutschland*. Frankfurt/M.: Campus-Verlag, 1995 (Studien zur Arbeitsmarktforschung Vol. 8).
- Pannenberg, M. *Zur Evaluation staatlicher Qualifizierungsmaßnahmen in Ostdeutschland: das Instrument Fortbildung und Umschulung (FuU)*. Halle: Institute for Economic Research, 1996 (Discussion paper, 38).
- Papalexandris, N.; Nikandrou, I. Benchmarking employee skills: results from best practice firms in Greece. *Journal of European Industrial Training*, 2000, Vol. 24, No 7, p. 391-402.
- Parkes, D. et al. *A cross country analysis of curricular reform in vocational education and training in Central and Eastern Europe: report*. European Training Foundation. Luxembourg: EUR-OP, 1999 (Integration of work and learning).
- Parsons, T. *The social system*. New York: The Free Press, 1951.
- Patterson, R.F.; Litt, D. (eds) *New Webster's Dictionary*. Miami, FL: P.S.I. and Associates, 1989.
- Patton, M.Q. *Utilization-focussed evaluation: the new century text*. Third Edition. London: Sage, 1997.
- Patton, M.Q. *Qualitative research and evaluation methods*. Third edition. London,: Sage, 2002.
- Pawson, R.; Tilley, N. *Realistic evaluation*. London: Sage, 1997.
- Payne, J. *England and Wales youth cohort study: options at 16 and outcomes at 24: a comparison of academic and vocational education and training routes*. Sheffield: DfEE, 1995.
- Payne, J. *Options at 16 and outcomes at 24. A comparison of academic and vocational education and training routes*. Sheffield: DfEE, 1997 (Youth cohort report, 35).
- Payne, J. et al. *Employment training and employment action: an evaluation by the matched comparison method*. Sheffield: Department for Education and Employment, 1996 (Research series, 74).
- Pazos, M.; Zapico-Goñi, E. Program evaluation in Spain: taking off at the edge of the twenty-first century? In: Furubo, J.E.; Rist, R.C.; Sandahl, R. (eds) *International atlas of evaluation*. London: Transaction Publishers, 2002, p. 291-306.
- Pedersen, P.; Dall Schmidt, T. *Search activity and re-employment – a European perspective*. Colchester: European Panel Analysis Group, 2002 (EPAG Working paper, 34).

- Picciotto, R. Evaluation in the World Bank. In: Furubo, J.E.; Rist, R.C.; Sandahl, R. (eds) *International atlas of evaluation*. London: Transaction Publishers, 2002, p. 425-437.
- Plewis, I.; Preston, J. *Evaluating the benefits of lifelong learning: a framework*. London: Institute of Education, 2001 (The wider benefits of learning papers, 2).
- Polachek, S.W.; Robst, J. Employee labor market information: comparing direct world of work measures of workers' knowledge to stochastic frontier estimates. *Labour Economics*, 1998, Vol. 5, Issue 2, p. 231-242.
- Pollitt, C. Evaluation in Europe: boom or bubble? *Evaluation*, 1998, Vol. 4, No 2, p. 214-224.
- Pont, B.; Werquin, P. Look, listen and learn: an international evaluation of adult learning. In: Descy, P.; Tessaring, M. (eds) *The foundations of evaluation and impact research. Third report on vocational training research in Europe: background report*. Luxembourg: EUR-OP, 2004 (Cedefop Reference series, 58).
- Power, M. *The audit society: rituals of verification*. Oxford: University Press, 1999.
- Prais, S.J. *Productivity, education and training: an international perspective*. Cambridge University Press, 1995.
- Preskill, H.; Torres, R.T. *Evaluative inquiry for learning in organizations*. Thousand Oaks, CA: Sage Publications, 1999.
- Preston, J.; Hammond, C. *The wider benefits of further education. Practitioner views*. London: Institute of Education, 2002.
- Prey, H. *Wirkungen von Maßnahmen staatlicher Arbeitsmarkt- und Beschäftigungspolitik*. Konstanz: University of Konstanz, 1997 (CILE Discussion paper, 45).
- Prey, H. *Wirkungen staatlicher Qualifizierungsmaßnahmen. Eine empirische Untersuchung für die Bundesrepublik Deutschland*. Stuttgart: Paul Haupt-Verlag, 1999.
- Pritchett, L. *Where has all the education gone?* Washington DC: The World Bank, 1995 (Policy research working paper, WPS 1581).
- Pukkinen, T.; Romijn, C.; Elson-Rogers, S. *Funding continuing training in small and medium-sized enterprises. Discussion and case studies from across the EU*. Cedefop Panorama series, 17. Luxembourg: EUR-OP, 2001.
- PURE. *Public funding and private returns to education. Final report*. 2001 (Project final report series – TSER Area 2 – SOE2-CT98-2044).
- Putnam, R. *Making democracy work: civic traditions in modern Italy*. Princeton: University Press, 1993a.
- Putnam, R. The prosperous community: social capital and public life. *The American Prospects*, 1993b, 13, p. 35-42.
- Putnam, R. Bowling alone: America's declining social capital. *Journal of Democracy*, 1995a, Vol. 6, Issue 1, p. 65-78.
- Putnam, R. *The case of missing social capital*. Working paper, Harvard University, 1995b.
- Putnam, R. *Bowling alone: the collapse and revival of American community*. New York: Simon and Schuster, 2000.
- Rabe-Kleberg, U. Frauen auf dem Weg zur Bildungsbiographie? *Frauenforschung*, 1994, Vol. 11, No 4, p. 5-16.
- Rantakeisu, U. et al. *Youth unemployment and social exclusion in Sweden*. In: Kieselbach, T. (ed.) *Youth unemployment and social exclusion: a comparison of six European countries*. Bremen: University of Bremen, 2000 (Psychology of social inequality series, Vol. 9, Yuseder Publication, No 1).
- Resnick, L. *Education and learning to think*. Washington, DC: National Academy Press, 1987.
- Rist, R.C.; Paliokas, K.L. The rise and fall (and rise again?) of the evaluation function in the U.S. government. In: Furubo, J.E.; Rist, R.C.; Sandahl, R. (eds) *International atlas of evaluation*. London: Transaction Publishers 2002, p. 225-245.
- Roberts, B. *Biographical research*, Buckingham: Open University Press, 2002.
- Röhrs, H. Georg Kerschensteiner (1852-1932). *Prospects: the Quarterly Review of Comparative Education*, 1993, Vol. 23, No 3 and 4, p. 807-822.
- Romer, P.M. Endogenous technological change. *Journal of Political Economy*, 1990, Vol. 98, Issue 5, p. S71-S102.

- Romijn, H.; Albaladejo, M. *Determinants of innovation capability in small UK firms: an empirical analysis*. Oxford: QEH – Queen Elizabeth House, 2000 (Working paper, 40).
- Rossi, P.H.; Freeman, H. *Evaluation: a systematic approach*. Fourth edition. Beverly Hills, CA: Sage Publications, 1988.
- Rossi, P.H.; Freeman, H.; Lipsey, M. *Evaluation: a systematic approach*. Sixth edition. London: Sage, 1999.
- Ruggles, R. The state of the notion: knowledge management in practice. *California Management Review*, 1998, Vol. 40, No 3, p. 80-89.
- Rychen, D.S. An overarching conceptual framework for assessing key competences in an international context. Lessons from an interdisciplinary and policy-oriented approach. In: Descy, P.; Tessaring, M. (eds) *The foundations of evaluation and impact research. Third report on vocational training research in Europe: background report*. Luxembourg: EUR-OP, 2004 (Cedefop Reference series, 58).
- Sackmann, R. Generationsspezifische Arbeitsmarktchancen im internationalen Vergleich. In: Berger, P. (ed.) *Die Erwerbsgesellschaft*. Opladen: Leske und Budrich, 2001.
- Sanderson, I. Evaluation in complex policy systems. *Evaluation: The International Journal of Theory, Research and Practice*, 2000, Vol. 6, Issue 4, p. 433-454.
- Sauer-Schiffer, U. *Biographie und Management. Eine qualitative Studie zum Leitungsverhalten von Frauen in der Erwachsenenbildung*. Münster: Waxmann, 2000.
- Schaeper, H.; Kühn, T.; Witzel, A. Diskontinuierliche Erwerbskarrieren und Berufswechsel in den 1990ern: Strukturmuster und biographische Umgangsweisen betrieblich ausgebildeter Fachkräfte. *Mitteilungen aus der Arbeitsmarkt- und Berufsforschung*, 2000, No 1, p. 80-99.
- Schallberger, U.; Spiess Huld, C. Die Zürcher Längsschnittstudie von der Schulzeit bis zum mittleren Erwachsenenalter. *Zeitschrift für Soziologie der Erziehung und Sozialisation*, 2001, Vol. 21, No 1, p. 80-89.
- Schmid, G. New public management of further training. In: Schmid, G.; O'Reilly, J.; Schömann, K. (eds) *International handbook of labour market policy and evaluation*. Cheltenham: Edward Elgar, 1996, p. 747-790.
- Schmid, G.; O'Reilly, J.; Schömann, K. (eds) *International handbook of labour market policy and evaluation*. London: Edward Elgar, 1996.
- Schmitt von Sydow, H. *White paper on European governance*. Report of the Working group 'Evaluation and Transparency'. European Commission, 2002.
- Schnabel, R.; Schnabel, I. *Family and gender still matter. The heterogeneity of returns to education in Germany*. Mannheim: ZEW, 2002 (Discussion paper, 02-67).
- Schömann, K. *The dynamics of labour earnings over the life course. A comparative and longitudinal analysis of Germany and Poland*. Berlin: Ed. Sigma, 1994.
- Schömann, K.; Becker, R. Participation in further education over the life course: a longitudinal study of three birth cohorts in the Federal Republic of Germany. *European Sociological Review*, 1995, Vol. 11, No 2, p. 187-208.
- Schrijvers, C.T.M. et al. De achtergronden van sociaal-economische gezondheidsverschillen: resultaten uit het GLOBE onderzoek. In: Stronks, K.; Hulshof, J. (eds) *De kloof verkleinen*. Assen: Koninklijke Van Gorcum 2001.
- Schuller, T. et al. *Learning, continuity and change in adult life*. London: Centre for Research on the Wider Benefits of Learning, 2002.
- Schumann, K.F.; Mariak, V. Benachteiligung Jugendlicher im Bildungssystem und auf dem Arbeitsmarkt als Weichenstellung für eine kriminelle Karriere – ein Mythos? In: *Das Jugendkriminalrecht als Erfüllungshilfe gesellschaftlicher Erwartungen?* Dokumentation des Symposiums an der kriminologischen Forschungsstelle der Universität zu Köln, 1995, p. 178-189.
- Schütz, H.; Speckesser, S.; Schmid, G. *Benchmarking labour market performance and labour market policies: theoretical foundations and applications*. Berlin: WZB –

- Wissenschaftszentrum für Sozialforschung, 1998 (Discussion paper, FS I 98–205).
- Schwandt, T. Evaluation as practical hermeneutics. *Evaluation: The International Journal of Theory, Research and Practice*, 1997, Vol. 3, No 1, p. 69-83.
- Scriven, M. *Handbook for model training program in qualitative educational evaluation*. Berkeley: University of California Press, 1973.
- Scriven, M. *Evaluation thesaurus*. Fourth edition. London: Sage, 1991.
- Seeber, S.; Krekel, E.M.; Buer, J.V. (eds) *Bildungscontrolling. Ansätze und kritische Diskussionen zur Effizienzsteigerung von Bildungsarbeit*. Frankfurt/M: Peter Lang, 2000.
- Shapiro, H.; Christensen, F. *Pædagogisk Grundlagsnotat, Arbejdsnotat*. Copenhagen: Ministry of Education, 1999.
- Shavit, Y.; Müller, W. *From school to work: a comparative study of educational qualifications and occupational destinations*. Oxford: Clarendon Press, 1998.
- Sianesi, B.; van Reenen, J. *The returns to education: a review of the empirical macro-economic literature*. London: IFS – Institute for Fiscal Studies, 2002 (IFS working papers, W02/05).
- Skocpol, T.; Fiorina, M. Making sense of the civic engagement debate. In: Skocpol, T.; Fiorina, M. (eds) *Civic engagement in American democracy*. Washington DC: Brookings Institution Press, 1999.
- Skov, P. *Ungees fremtid – meget afgøres tidligt*. Copenhagen: Danmarks Pædagogiske Institut – DPI, 1998.
- Smith, M.L. The whole is greater: combining qualitative and quantitative approaches in evaluation studies. *New Directions in Program Evaluation*, 1986, p. 37-54.
- Smith, M.L.; Glass, G.V. *Research and evaluation in education and the social sciences*. Englewood Cliffs, NJ: Prentice Hall, 1987.
- Sofer, C. (ed.) *Schooling, training and transitions: an economic perspective (STT)*. Orléans: University of Orléans, 2000 (Final report).
- Sokou, K. et al. Youth unemployment and social exclusion in Greece. In: Kieselbach, T. (ed.) *Youth unemployment and social exclusion: a comparison of six European countries*. Bremen: University of Bremen, 2000 (Psychology of social inequality series, Vol. 9, Yuseder Publication, No 1).
- Solga, H. 'Low education' as a social stigma. A sociological explanation for the decreasing employment opportunities of the low educated. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 2002, Vol. 54, p. 476-505.
- Solomon, L.C. The relation between schooling and savings behavior: an example of the indirect effects of education. In: Juster, F.T. (eds) *Education, income, and human behavior*. New York: McGraw-Hill, 1975.
- Solow, R.M. Technical change and aggregate production function. *Review of Economics and Statistics*, 1957, Vol. 39, p. 312-320.
- Staat, M. *Empirische Evaluation von Fortbildung und Umschulung – Schriftenreihe des ZEW 21*. Baden-Baden: Nomos Verlagsgesellschaft, 1997.
- Stame, N. Evaluation in Italy: an inverted sequence from performance management to program evaluation. In: Furubo, J.E.; Rist R.C.; Sandahl, R. (eds) *International atlas of evaluation*. London: Transaction Publishers, 2002, p. 273-291.
- Starrin, B. et al. International debate on social exclusion. In: Kieselbach, T. (ed.) *Youth unemployment and social exclusion: a comparison of six European countries*. Bremen: University of Bremen, 2000 (Psychology of social inequality series, Vol. 9, Yuseder Publication, No 1).
- Steiner, V.; Lauer, C. Private Erträge von Bildungsinvestitionen in Deutschland [Private returns to education in Germany]. *Beihefte zur Konjunkturpolitik. [Applied Economics Quarterly]*, Vol. 51, 2000, p. 71-101.
- Stern, E. Philosophies and types of evaluation research. In: Descy, P.; Tessaring, M. (eds) *The foundations of evaluation and impact research. Third report on vocational training research in Europe: background report*. Luxembourg: EUR-OP, 2004 (Cedefop Reference series, 58).

- Stockmann, R. *The sustainability of development cooperation*. Baden-Baden: Nomos Verlag, 1997.
- Stockmann, R. (ed.) *Evaluationsforschung: Grundlagen und ausgewählte Forschungsfelder*. Opladen: Leske und Budrich, 2000a (Sozialwissenschaftliche Evaluationsforschung, 01).
- Stockmann, R. Evaluation in Deutschland. In: Stockmann, R. (ed.) *Evaluationsforschung. Grundlagen und ausgewählte Forschungsfelder*. Opladen: Leske und Budrich, 2000b, p. 11-40 (Sozialwissenschaftliche Evaluationsforschung, 01).
- Stone, V.; Cotton, D.; Thomas, A. *Mapping troubled lives: young people not in education, employment or training*. London: DfEE – Departement for Employment and Education, 2000 (Research Brief, 181).
- Straka, G.A. Measurement and evaluation of competence. In: Descy, P.; Tessaring, M. (eds) *The foundations of evaluation and impact research. Third report on vocational training research in Europe: background report*. Luxembourg: EUR-OP, 2004 (Cedefop Reference series, 58).
- Stronach, I.; Morris, B. Polemical notes on educational evaluation in the age of 'policy hysteria'. *Evaluation and Research in Evaluation*, 1994, Vol. 8, p. 1-15.
- Struyven, L.; Steurs, G. Quasi-market reforms in employment and training services: first experiences and evaluation results. In: Descy, P.; Tessaring, M. (eds) *Evaluation of systems and programmes. Third report on vocational training research in Europe: background report*. Luxembourg: EUR-OP, 2004 (Cedefop Reference series, 57).
- Stufflebeam, D.L. Foundational models for 21st century program evaluation. In: Stufflebeam, D.L.; Madaus, G.F.; Kellaghan, T. *Evaluation models: viewpoints on educational and human services evaluation*. Second edition. Dordrecht: Kluwer Academic Publisher, 2000.
- Summa, H., Toulemonde, J. Evaluation in the European Union: addressing complexity and ambiguity. In: Furubo, J.E.; Rist, R.C.; Sandahl, R. (eds). *International atlas of evaluation*. London: Transaction Publishers, 2002, p. 407-424.
- Summers, R.; Heston, A. The Penn world table (Mark 5): an expanded set of international comparisons, 1950-1988. *Quarterly Journal of Economics*, 1991, Vol. 106, Issue 2, p. 327-368.
- Swan, T.W. Economic growth and capital accumulation. *Economic Record*, 1956, Vol. 32, p. 334-361.
- Temple, J. *Education and economic growth*. Paper presented at the HM Treasury seminar on economic growth and government policy, 12 October 2000.
- Temple, J. Growth effects of education and social capital in OECD countries. *OECD Economic Studies*, 2001, No 33, Vol. 2, p. 57-101.
- Tessaring, M. (ed.) *Vocational education and training – the European research field*. Background report. Luxembourg: EUR-OP, 1998 (Cedefop Reference document).
- Tessaring, M. *Training for a changing society: a report on current vocational education and training research in Europe*. Second edition. Luxembourg: EUR-OP, 1999 (Cedefop Reference series).
- Tessaring, M. *Achieving the Lisbon goals: implications for human resources investment in education and training and economic development*. Thessaloniki: Cedefop, 2003 (Working paper).
- The Economist. *Business database dictionary*. Available from Internet: <http://www.economist.com/businessDatabase/Dictionary.cfm?id=ID8D747D13-8E5B-11D5-8499-00508B2CCA66> [cited 9.9.2004].
- Tissot, P. *Terminology of vocational training policy: a multilingual glossary for an enlarged Europe*. Luxembourg: EUR-OP, 2004.
- Toivanen, O.; Stoneman, P.; Bosworth, D. *Innovation and the market value of UK firms, 1989-95*. Coventry: Warwick Business School, 1998 (Working paper).
- Torney-Purta, J. et al. *Citizenship and education in twenty-eight countries. Civic knowledge and engagement at age fourteen*. Amsterdam: International Association for the Evaluation of Educational Achievement, 2001.

- Torp, H. et al. The first Norwegian experiment. In: Jensen, K.; Madsen, P.K. (eds) *Measuring labour market measures: evaluating the effects of active labour market policy initiatives*. Copenhagen: Ministry of Labour, 1993, p. 97-140.
- Toulemonde, J. Evaluation culture(s) in Europe: differences and convergence between national practices. *DIW Vierteljahresheft*, 2000, No 3, p. 350-357.
- Trostel, P.; Walker, I.; Woolley, P. Estimates of the economic return to schooling for 28 countries. *Labour Economics*, 2002, Vol. 9, No 1, p. 1-16.
- Trouvé, P. The employment and training practices of SMEs: examination of research in five EU Member States. In: Descy, P.; Tessaring, M. (eds) *Training in Europe: second report on vocational training research in Europe 2000: background report*. Luxembourg: EUR-OP, 2001, Vol. 2, p. 91-232 (Cedefop Reference series, 3008).
- Tsakoglou, P. Aspects of inequality in Greece: measurement, decomposition and inter-temporal change: 1974, 1982. *Journal of Development Economics*, 1993, Vol. 40, p. 53-74.
- Tsakoglou, P. Changes in inequality in Greece in the 1970s and the 1980s. In: Gottschalk, P.; Gustafsson, B.; Palmer, E. (eds) *Changing patterns in the distribution of economic welfare: what happened during the 1980s?* Cambridge: University Press, 1997.
- Tuma, N.B. Effects of labour market structure on job-shift patterns. In: Heckman, J.J.; Singer, S. (eds) *Longitudinal analysis of labour market data*. Cambridge: University Press, 1985.
- Tuominen, E. *The active years of the life course and pension provision in transition. Changes in the duration of employment during the life span in recent decades. The Finish example*. The Year 2000 International Research Conference on Social Security, Helsinki, 25-27 September 2000. Helsinki: ISSA – International Social Security Association, 2000. Available from Internet: <http://www.issa.int/pdf/helsinki2000/topic2/2Tuominen.PDF> [cited 2.9.2004].
- Unwin, L.; Wellington, J. *Young people's perspectives on education, training and employment*. London: Kogan Page 2001.
- US Department of Education. *Systemic reform in the professionalism of educators*, 1995. Available from Internet: <http://www.ed.gov/pubs/SER/ProfEd/index.html> [cited 9.9.2004].
- Valovirta, V. Evaluation utilization as argumentation. *Evaluation – the International Journal of Theory, Research and Practice*, 2002, Vol. 8, No 1, p. 60-80.
- van de Werfhorst, H.G. *Systems of educational specialization and labour market outcomes in Norway, Australia, and the Netherlands*. Oxford: Nutfield College, 2002. Available from Internet: http://www.nuff.ox.ac.uk/users/werfhorst/ED-LM_NOAUSNL.pdf [cited 2.9.2004].
- van der Knaap, P. Policy evaluation and learning: feedback, enlightenment, or argumentation. *Evaluation: The International Journal of Theory, Research and Practice*, 1995, Vol. 1, No 2, p. 189-216.
- Vedung, E. Evaluation research and fundamental research. In: Stockmann, R. (ed.) *Evaluationsforschung: Grundlagen und ausgewählte Forschungsfelder*. Opladen: Leske und Budrich, 2000a (Sozialwissenschaftliche Evaluationsforschung, 01).
- Vedung, E. *Public policy and programme evaluation*. New Brunswick, NJ: Transaction Publishers, 2000b.
- Veierød, M.B.; Thelle, D.S.; Laake, P. Diet and risk of cutaneous malignant melanoma: a prospective study of 50 757 Norwegian men and women. *International Journal of Cancer*, 1997, Vol. 71, p. 600-604.
- Veum, J.R. *Training, wages and the human capital model*. Washington, DC: US Bureau of Labor Statistics, 1995 (NLS report, 96-31).
- Viertel, E. et al. From project to policy evaluation in vocational education and training – possible concepts and tools. Evidence from countries in transition. In: Descy, P.; Tessaring, M. (eds) *The foundations of evaluation and impact research. Third report on vocational training research in Europe: background report*. Luxembourg: EUR-OP, 2004 (Cedefop Reference series, 58).

- von Bertalanffy, L. An outline of general systems theory. *British Journal for the Philosophy of Science*, 1950, Vol. 1, No 2, p. 139-164.
- Walsh, K.; Parsons, D.J. Active policies and measures: impact on integration and reintegration in the labour market and social life. In: Descy, P.; Tessaring, M. (eds) *Impact of education and training. Third report on vocational training research in Europe: background report*. Luxembourg: EUR-OP, 2004 (Cedefop Reference series, 54).
- Watts, M. Aggressive youth cultures and hate crime: skinheads and xenophobic youth in Germany. *American Behavioral Scientist*, 2001, Vol. 45, No 4, p. 600-615.
- Weinert, F.E. Concepts of competence: a conceptual clarification. In: Rychen, D.S.; Salganik, L.H. (eds) *Defining and selecting key competencies*. Göttingen: Hogrefe and Huber, 2001, p. 45-66.
- Weiss, C.H. Using research in the policy process: potential and constraints. *Policy Studies Journal*, 1976, Vol. 4, p. 224-228.
- Weiss, C.H. Nothing as practical as good theory: exploring theory-based evaluation for comprehensive community initiatives for children and families. In: Connell, J.P.; et al. (eds) *New approaches to evaluating community initiatives: concepts, methods and contexts*. New York: The Aspen Institute, 1995.
- West, A.; Sparkes, J.; Balanov, T. *Demand-side financing: a focus on vouchers in post-compulsory education and training. A discussion paper and case studies*. Luxembourg: EUR-OP, 2000 (Cedefop Dossier, 6003).
- Westphalen, S.Å. Reporting on human capital: objectives and trends. In: Descy, P.; Tessaring, M. (eds) *Training in Europe. Second report on vocational training research in Europe 2000: background report*. Luxembourg: EUR-OP, 2001, Vol. 2, Part 4, p. 249-278 (Cedefop Reference series, 3008).
- White, M.; Lakey, J. *The restart effect: evaluation of a labour market programme for unemployed people*. London: Policy Studies Institute, 1992.
- Wholey, J.S. Evaluierung – Grundlage und Voraussetzung für leistungsfähige Programme. In: Hellstern, G.-M.; Wollmann, H. (eds) *Handbuch zur Evaluierungsforschung*. Vol. 1. Opladen: Westdeutscher Verlag, 1984.
- Wilson, R.A.; Briscoe, G. The impact of human capital on economic growth: a review. In: Descy, P.; Tessaring, M. (eds) *Impact of education and training. Third report on vocational training research in Europe: background report*. Luxembourg: EUR-OP, 2004 (Cedefop Reference series, 54).
- Windham, D.M. *The OECD thematic reviews of education policy in South Eastern Europe: a proposed outline for authors of country chapters*. Paris: OECD, 2000 (unpublished draft).
- Winter-Ebmer, R. *Evaluating an innovative redundancy-retraining project: the Austrian Steel Foundation*. London: CEPR – Centre for Economic policy research, 2001 (CEPR discussion paper DP2776).
- Wolbers, M.H. The effects of level of education on mobility between employment and unemployment in the Netherlands. *European Sociological Review*, 2000, Vol. 16, No 2, p. 185-200.
- Wolf, A. *Does education matter? Myths about education and economic growth*. London: Penguin Group, 2002.
- Wolfe, B.; Zuveska, S. Non-market outcomes of schooling. *International Journal of Education Research*, 1997, Vol. 27, No 6, p. 491-502.
- Woolcock, M. Social capital and economic development: toward a theoretical synthesis and policy framework. *Theory and Society*, 1998, Vol. 27, Issue 2, p. 151-208.

- Woolcock, M. *The place of social capital in understanding social and economic outcomes*. Paper presented at the international symposium on the contribution of human and social capital to sustained economic growth and well-being, organised by the OECD and HRDC, Quebec City, Canada, 19-21 March 2000.
- Wooley, F. *Social cohesion and voluntary activity: making connections*. Paper presented at the CSLS Conference on *the State of Living Standards and the quality of life in Canada*, held in Ottawa, Ontario on 30-31 October 1998.
- World Bank (2002). Social capital home page: www.worldbank.org/poverty/scapital/index.htm [cited 13.9.2004].
- Worthen, B.R.; Sanders, J.R.; Fitzpatrick, J.L. *Program evaluation: alternative approaches and practical guidelines*. New York: Longman, 1997.
- Yuseder-report. *Youth unemployment and social exclusion: a comparison of six European countries*. Opladen: Leske und Budrich, 2000 (Psychology of social inequality series, Vol. 9, Yuseder Publication, No 2).

Cedefop (European Centre for the Development of Vocational Training)

The value of learning Evaluation and impact of education and training

Third report on vocational
training research in Europe:
synthesis report

Pascaline Descy
Manfred Tessaring

Luxembourg:
Office for Official Publications of the European Communities

2005 – VI, 323 pp. – 21 x 29.7 cm
(Cedefop Reference series; 61 – ISSN 1608-7089)

ISBN 92-896-0336-4

Cat. No: TI-64-04-256-EN-C

Price (excluding VAT) in Luxembourg: EUR 40

No of publication: 3042 EN

Cedefop's third research report *The value of learning: evaluation and impact of education and training* addresses the contribution of education and training to realising a knowledge-based society. It specifies the material and non-material benefits of learning for individuals, enterprises and society as a whole. It also discusses ways of improving the design and implementation of education and training by purposeful evaluation.

The report allows the reader to gain an overview of:

- the main functions and uses of evaluation, their philosophies and theories;
- quantitative methods for evaluation and impact research at micro and macro level as well as methods of qualitative inquiry and their application in evaluation;
- tools and approaches for evaluating and comparing education and training systems, for accompanying the implementation – and assessing the outcomes – of programmes and reforms, illustrated by case studies;
- impact research, quantifying the material and non-material benefits of education, training and skills at macro level of nations and regions, their contribution to enterprise performance and their influence on individuals' biographies and careers.

Pascaline Descy, Manfred Tessaring

EN

The value of learning: evaluation and impact of education and training

**Third report on vocational
training research in Europe**
Synthesis report



European Centre for the
Development of Vocational Training

Europe 123, GR-570 01 Thessaloniki (Pylea)
Postal address: PO Box 22427, GR-551 02 Thessaloniki
Tel. (30) 23 10 49 01 11, Fax (30) 23 10 49 00 20
E-mail: info@cedefop.eu.int
Homepage: www.cedefop.eu.int
Interactive website: www.trainingvillage.gr

Price (excluding VAT) in Luxembourg: EUR 40

3042 EN

