

EN

Evaluation of systems and programmes

Third report on vocational
training research in Europe
Background report



Evaluation of systems and programmes

Third report on vocational training research in Europe: background report

Pascaline Descy,
Manfred Tessaring (eds)

Cedefop Reference series; 57

Luxembourg: Office for Official Publications of the European Communities, 2004



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, 2004

ISBN 92-896-0279-1

© European Centre for the Development of Vocational Training, 2004
All rights reserved.

Printed in Italy

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-57001 Thessaloniki (Pylea)

Postal Address:
PO Box 22427
GR-55102 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

Edited by:

Cedefop

Pascaline Descy, Manfred Tessaring, Project managers

Published under the responsibility of:

Johan van Rens, Director

Stavros Stavrou, Deputy Director

Content

Preface	7
Evaluating the impact of reforms of vocational education and training: examples of practice <i>Mike Coles</i>	9
Evaluating systems reform in vocational education and training: learning from Danish and Dutch cases <i>Loek Nieuwenhuis, Hanne Shapiro</i>	53
Evaluation of EU and international programmes and initiatives promoting mobility: selected case studies <i>Wolfgang Hellwig, Uwe Lauterbach, Hermann-Günter Hesse, Sabine Fabriz</i>	93
Consultancy for free? Evaluation practice and culture in the European Union and central and eastern Europe. Findings from selected EU programmes <i>Bernd Baumgartl, Olga Strietska-Ilina, Gerhard Schaumberger</i>	161
Quasi-market reforms in employment and training services: first experiences and evaluation results <i>Ludo Struyven, Geert Steurs</i>	219
Evaluation activities in the European Commission <i>Josep Molsosa</i>	273
List of authors	288

Other volumes of the background report

The background report to the third Research report is composed of two other volumes published separately the content of which is detailed below.

The foundations of evaluation and impact research

Descy P., Tessaring M. (eds)

Preface

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 Radić

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

Impact of education and training

Descy P., Tessaring M. (eds)

Preface

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, 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

Preface

The series of reports on vocational education and training (VET) research have been published by Cedefop since 1998 ⁽¹⁾. The reports provide a comprehensive review of current research in initial and continuing VET in Europe, research 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.

Each research report consists of a background report of several volumes with contributions from renowned researchers (this publication) and a synthesis report elaborated by Cedefop experts.

Third research report: evaluation and impact of education and training

This third report informs on current research on evaluation and the impact of education and training on individuals, enterprises and society and economy in general, including:

- (a) the assessment of education and training systems;
- (b) the implementation and outcomes of programmes and reforms with a VET component;
- (c) impact and cost-benefit research, quantifying the contribution of education, training and skills on, for example, earnings, economic growth, employment and social inclusion.

The report serves both to inform and improve policy and practice, and further develop research. It also contributes to the discussion on the overall European goals expressed by the European Council at its Lisbon and follow-up summits.

Demonstrating the contribution of education and training towards realising a knowledge-based society and specifying the diverse benefits at all levels as done by 'summative evaluation' or impact research – is as important as indicating ways to improve the design and implementation

of education and training programmes or measures by 'formative evaluation'.

The background report

The background report gathers contributions from renowned experts and researchers. They allow the reader to approach evaluation and impact research from various angles: individual, enterprise and macro-system level by also considering essential basics on the foundation, approaches, standards – and limitations – of evaluation and impact research.

Contributions have been regrouped into three broad themes, published in separate volumes:

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

Evaluation of systems and programmes

The present volume discusses evaluations of education and training systems and programmes. It gathers case studies that provide an insight into recent evaluation work, its methods and results.

The contribution of M. Coles looks at current practices to assess the impact of education and training reforms in various European countries. By analysing evaluative reports, they provide insights into the ways evaluations have been conducted, their role in the reform process and the kind of impacts the reforms have produced. L. Nieuwenhuis and H. Shapiro complete the picture by looking in depth and comparing the evaluations carried out to assess and accompany the VET system reforms in Denmark and in the Netherlands.

W. Hellwig et al. analyse and look critically at evaluations of mobility and exchange programmes implemented either with the support of the European Union, such as the mobility measure or the Community study visit programme of Leonardo da Vinci, or in binational

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

exchanges, such as the German-French youth foundation.

B. Baumgartl et al. present case studies of evaluations of the NARIC network in the Socrates programme and Phare VET programmes in Bulgaria, the Czech Republic and Slovakia and Leonardo da Vinci in the Czech Republic.

The impact of the introduction of market forces in employment and training services for integrating job seekers is the subject of the contribution of L. Struyven and G. Steurs. Finally, J. Molsosa takes stock and proposes future steps for evaluating activities in the European Commission.

**Stavros
Stavrou**

**Manfred
Tessaring**

**Pascaline
Descy**

References

- Cedefop. *Vocational education and training – the European research field: background report* (2 volumes). Luxembourg: Office for Official Publications of the European Communities, 1998 (Cedefop Reference series).
- Descy, P.; Tessaring, M. (eds) *Training and learning for competence: second report on vocational training research in Europe: synthesis report*. Luxembourg: Office for Official Publications of the European Communities, 2001a (Cedefop Reference series, 6).
- Descy, P.; Tessaring, M. (eds) *Training in Europe: second report on vocational training research in Europe 2000: background report* (3 volumes). Luxembourg: Office for Official Publications of the European Communities, 2001b (Cedefop Reference series, 3008).
- Tessaring, M. *Training for a changing society: a report on current vocational education and training research in Europe* (second edition). Luxembourg: Office for Official Publications of the European Communities, 1999 (Cedefop Reference series).



Evaluating the impact of reforms of vocational education and training: examples of practice

Mike Coles

Abstract

In this study 19 evaluative reports of national reforms to vocational education and training (VET) across 12 European Union countries have been selected and analysed. Criteria have been developed which clarifies the analytical framework. The study of these reports of reforms provides insights into the ways evaluations have been conducted, the role of evaluations in the reform process and the kinds of impacts the reforms have made.

A range of large-scale national reforms to VET is examined covering initial vocational training (IVT), continuing vocational training (CVT), VET in higher education, VET for employability, qualification development and financing. The trend away from central control of VET planning and financing is evident as is a deepening appreciation of the potential of VET for broadening the school curriculum and for improving the life chances of individuals.

Countries differ greatly in the ways they use evaluation. In some countries evaluation appears not to be used as a development tool and in others it is developed simultaneously with the reform process. Published impact studies are rare; the most common form of evaluation programme is a mid-term and/or summative stakeholder survey. However, there are some good examples of published impact analysis and the methodological rigour in these can be used to develop confidence in tackling impact studies. The study offers evidence-based guidance on making evaluation programmes more scientific and more impact oriented.

Table of contents

1. Introduction	12
1.1. The nature of VET	12
1.2. Some definitions and limitations	13
2. Selection of reports for analysis	14
3. Criteria for judging reforms	18
4. Evaluating VET reform	19
4.1. VET reform	19
4.2. Evaluation	21
4.3. Context of reform	24
4.4. Goals	24
4.5. Planning of reforms	25
4.6. Management of reforms	25
4.7. Drivers	26
4.8. Impact	26
5. Financing VET: reforms and evaluation	27
5.1. Financing reforms	27
5.2. Evaluation of financing	28
6. Levels of evaluation	29
6.1. Establishing a baseline	29
6.2. Impact and timing of measurement	30
6.3. The individual	30
6.4. The organisation	34
6.5. The aggregate level of the economy and society	37
7. Towards better evaluation of reforms	40
8. Summary and conclusions	42
List of abbreviations	44
Annex: common evaluation criteria	45
References	49

List of tables

Tables

Table 1: Reports used as the evidence base	15
Table 2: Classification of reforms to learning and learning environments	21
Table 3: Broad review of VET reforms and evaluation methods in this study	23
Table 4: Measuring the impact of new funding arrangements	28
Table 5: Proportion of Fachhochschule leavers who feel disadvantaged by their counterparts in 'normal universities' with regard to selected aspects	31
Table 6: How much an apprenticeship course is helping to...	31
Table 7: Increases in salary from the first employment in 1992 to employment in 1997	31
Table 8: The measures of success for two active youth labour market programmes	32
Table 9: Why a modern apprenticeship is attractive	32
Table 10: Motives for abandoning apprenticeship	33
Table 11: Effect of training on long-term unemployed	33
Table 12: Changes in workers as a result of VET	35
Table 13: Employers' views on the reform	35
Table 14: <i>Fachhochschule</i> graduates compared to graduates from traditional universities	36
Table 15: Are you interested in employing more graduates from <i>Fachhochschule</i> in the future?	37
Table 16: Five years after gaining a qualification how many young people in France are registered as unemployed or inactive?	38
Table 17: Destinations on completion of Portuguese IVT programme	38
Table 18: Labour-market situation immediately following VTOS by certification achievements	38

1. Introduction

The main objective of this study is to provide a critical review of research dealing with evaluation of education and training system reforms that influence the organisation and outcomes of education and training at macro level. To develop this review a selection of national reforms to vocational education and training (VET) has been made and reports on these reforms analysed against criteria to provide insights into the ways evaluations have been conducted, the role of evaluations in programmes of national reforms and the impacts these reforms have made.

The task of looking at national reforms and finding evaluative reports on them has been a complex task. The specific nature of this study has meant that many reports fail to meet the exacting specification required and consequently researching sources of potentially useful information has been time consuming. In the end 19 reports have been identified as relevant to the study. Ten are what might be termed 'core' reports as they are especially rich in data on a reform, its evaluation process and the evaluation of the effects of the reform. Two of the 19 reports (from Denmark and the Netherlands) are specific case studies in Part IV of the third research report, but they are not considered in great detail in this study.

1.1. The nature of VET

VET is a complex subject of research because it is so wide ranging, covering multiple forms of activity from parts of mainstream initial education to short-term, highly focused job specific skill development. There are also multiple layers of activity on where learning takes place, who organises it, who pays for it and who certifies it. The value of VET is located with individuals, communities, institutions, trade unions, businesses and governments at regional, national and international levels. Development and reform of VET is therefore subject to a multitude of influences, controls and incentives. Setting up a large-scale reform to a country's VET system

demands high levels of consultation, commitment from national agencies and detailed planning. There are clearly many ways to manage a national reform ranging from a carefully conceived, multistage plan to one that sets out to achieve aims through multiple activities. In the latter type, actions are often dependent on responses to recent changes in the system and so method corresponds, for example, to setting in place an integrated transport system where the development of the train network depends on the functioning of other means of transport and the choices people make as different ways of travelling become available.

Evaluation processes also take on different roles in these different styles of reform. Some represent a feedback system to central planners on progress and are mechanical in nature using targets or performance indicators as accountability tools. We could label these evaluations as 'summative'. 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. This style could be labelled 'formative'. The complex nature of VET and its social embeddedness makes a single prestructured summative evaluation of limited value. The evaluation process might be considered to be about learning about VET change itself and a formative evaluation process is, perhaps, essential. In fact, drawing on the experience of researching the evidence for this study evaluations of national reforms are rarely singularly summative or singularly formative but fall within a spectrum of types ranging from one to the other.

All evaluation studies respond to their objectives and report on these. However some go further and report on the implications of findings for the system being evaluated. This extension can, in some reports, lead to substantial statements that have implications for reform managers. This 'added value' to evaluation

reports is generally associated with evaluations of the embedded or formative kind.

In this study a wide range of types of VET reform and evaluation processes have been encountered. The spectrum of types of reform and evaluation processes makes the analysis presented a useful illuminative tool for VET researchers and offers ideas for further, deeper analysis.

1.2. Some definitions and limitations

It is useful to state how certain key concepts are used in this report.

VET is defined as all organised forms of initial and continuing or further vocational education and training, independent of location, age of participants and their level of qualification. VET is to prepare for a certain occupation or employment and could also include an element of general education, for example the development of basic skills.

Reforms of VET are limited in this study to those that are national or system-wide. Reforms need not cover the whole VET system in a country to be included, indeed no examples of this root and branch scale of reform were found, however reforms were required to cover a substantial component of the VET system. Regional reforms have also been included when the results of the reform are likely to impact on

national provision. A distinct part of the focus on VET reform is to consider the financing measures. Some VET reforms include reform of financing arrangements as well as the content and delivery of VET. Some evaluation reports are not associated with any single reform. They are an evaluation of VET provision in a country from the perspective of a particular set of users. For example, evaluation of effects of qualifications on a full cohort of school leavers in 1992 has been carried out in France. Leaving out such studies would have reduced the quality of the analysis of types of evaluation.

Evaluation is research that involves use of social and economic methodologies to assess and improve the planning, implementation, monitoring, effectiveness and efficiency of measures or programmes. In this study we are concerned with evaluation of impact. We are looking for some idea of how changes to individuals, institutions, communities and systems are dependent on interventions embedded in the reform process. Stern (2003) refers to these evaluations as those that bring 'accountability for policy-makers' and distinguishes them from evaluations which aim to develop a reform programme, create knowledge about the VET system or aim for social improvement and change. The focus of these accountability evaluations is impact, outcomes and achievement of targets and value for money.

The next chapter outlines the research, giving an indication of scope, the work undertaken and the timescale involved.

2. Selection of reports for analysis

Searches for reports were protracted and often led to an inadequate description of the evaluation process, focusing instead on detailed descriptions of the planning and introduction of reforms. For this research there are three points of interest – the reform itself, the evaluation of the reform (process and outcomes) and how the reform and the evaluation relate to each other. The first point is always present, the second sometimes present and the latter is invariably missing. The intersection of the reform and the evaluation is important for two main reasons:

- (a) it gives an insight into the way in which the reform and the evaluation methodology are matched to each other by design. For example, how well the objectives for evaluation correspond with the goals of the reform;
- (b) it tells us how impact is being monitored and then weighed against other contextual factors.

Compiling a group of reports representative of the range available proved difficult; it was not possible to be as selective as one would have wished as the range of reports available was limited. Selection of reports was initially based on:

- (a) the significance of the reform in the country concerned;
- (b) the breadth and depth of the evaluation;
- (c) the extent of the focus on evaluation of impact.

In practice, the first goal was to find at least one report from each of the 15 EU nations. It was desirable to have good coverage of cultures and VET systems across countries and in some countries there was more than one report eligible for inclusion.

Exchanges with researchers have suggested that full reports of evaluations of VET reforms are not published as frequently as one might expect. Many reasons are cited by experts including a wish for confidentiality by funding agencies, lack of political will to publish challenging evaluations and the fairly common 'light touch' style of evaluation that yield low levels of analytical information. In the initial stages of the work over

200 references to potentially useful sources of evaluative information were identified. These were scrutinised against the broad requirements of this study. Essentially reports were required that:

- (a) had the subject of a major reform to VET within a country;
- (b) focused on a reform that was not EU funded (covered in other contributions to the third research report);
- (c) was written and published authoritatively (by high status people/organisations);
- (d) was evaluative in nature rather than descriptive;
- (e) focused on the impact of reforms;
- (f) included reference to ways of financing VET provision.

There were many promising references to reports that appeared to meet these requirements but as the reports were scrutinised in detail many fell away. This is a common pattern in metastudies; in the major ones into specific aspects of education in the UK ⁽¹⁾ it is not uncommon for 2 000 references to be reduced to 20 or so useful reports. The filtering process in this study showed that there are many evaluation reports but:

- (a) most are about the way a project worked rather than the impact a reform has had;
- (b) many are linked to EU programmes and not reforms initiated and funded by countries to meet national goals;
- (c) many are more concerned with regional or fragments of the VET system than system-wide reforms;
- (d) many are linked to broad initial education rather than vocational workforce development;
- (e) few have specific references to financing.

This paucity of outcome-based reporting is highlighted in evaluation literature, Griffith (2001) states in his review of school-to-work transitions most publications quote such variables as numbers of participants and numbers of work placements and few chart progress through to

(1) EPPI – Evidence based policy and practice – University of London, Institute of Education

outcome measures. Baker and Taylor (1998) report that the shortage of empirical research more generally is the issue; they found only 12 such studies published between 1983 and 1996 and these were small scale and limited in various ways.

The first sift of reports that related to all the key words (evaluation, impact, national reform, vocational, education, training) reduced the 200+ potentially useful reports to 33. These covered all EU countries and most aspects of VET and financing.

In some countries it is difficult to isolate a specific thread of evaluation on programmes. It is possible to take two fairly extreme views of this. Either there is no discernible, independent evaluation or, development and evaluation are so intertwined it is not possible to distinguish them. In the latter case development or reform takes place in relatively small stages. It would be useful to research reports of evolution of a system and trace the drivers of change. In so doing the process of evaluating current provision and making incremental changes as a result should become clear.

Some interesting reports were identified that were not conventional VET research publications but nevertheless had potential to contribute to this study. For example a Cedefop review paper (Richini, 2002) on steps to enhance lifelong learning in Italy proved useful for information on ways of better coordinating VET, two French

documents (Klein, 2002; Thireau, 2001) detailed how a single cohort of school leavers fared in the labour market and in further education and training. Another report of this type compared two UK methods of financing training (Felstead and Unwin, 1999). Two Swedish active labour market policies are systematically compared in another report (Larsson, 2000). These reports have been used alongside others in forming the evidence base for this study.

In the early stages of this work potential reforms were to be classified as core, suitable for indepth analysis, and associate, useful for supplementary material for illuminating, supporting, challenging and cross referencing findings about theories, approaches, methods and results in the core metaanalysis. This categorisation proved untenable in the light of difficulties in finding reports and, more importantly, in trying to cover types of reform and evaluation within six to eight reports of reforms. A decision was made to reduce the set of 33 reports to a single set of reports on 19 reforms and to treat them in the same way. Table 1 lists these and includes a justification for including the report in the study.

Having isolated the key reforms and their associated evaluative reports, each report was analysed against the criteria described earlier and is attached as an annex to this report. These criteria are discussed in the next chapter.

Table 1: **Reports used as the evidence base**

Country	Reform	Report reference	Justification for inclusion
Austria	<i>Die ersten FachhochschulAbsolventInnen am Arbeitsmarkt</i>	Wimmer, 2000	The development of <i>Fachhoch</i> schools provides a tertiary education focus. Students and enterprises were involved in the evaluation process.
Denmark	Reform 2000	Shapiro et al., 2001 Kaasbøl et al., 2001 Nielsen, 1997 Nieuwenhuis and Shapiro, 2003	This reform is included as it is a case study in the third research report. The reform is nationwide and involved a pilot phase and multiple evaluations. It also has a theoretical basis.
Finland	<i>Valtioneuvooston selonteko eduskunnalle uuisen koululakien vaikutuksista ja laeissa asetettujen tavoitteiden toteutumisesta</i>	Finnish Ministry of Education, 2000	This is a major reform to secondary education and adult education. There are changes to qualifications and new models of financing. The official evaluation report shows how evaluation is integrated with the reform process.

France	<i>Les bénéficiaires du contrat de qualification entrés à l'automne 1994: les enseignements de 4 vagues de panel</i>	Klein, 2002	This is an example of a longitudinal study with focus on the effects of a vocational qualification. A wide range of impact measures is covered.
Germany	<i>Personalqualifizierung in den neuen Bundesländern</i>	Neubert and Steinborn, 1999	It is difficult to identify a specific reform that has been evaluated in German VET. Recent initiatives may provide such reports. This report contains information about the impact of training and contains a financial dimension.
Greece	Evaluation of the labour-market policies and assessment of the influence of European employment strategy in Greece during the period 1997 to 2001.	VFA, 2002	While there is no specific VET reform in this report it is possible to study the effects of VET scientifically: case/control methodology was used on a large sample. The case group and the control group were strikingly well matched demographically.
Ireland	Leaving certificate (applied)	Irish Ministry of Education, 2002	This is a large-scale qualification development with a strong VET dimension. A strong decentralised evaluation programme accompanies it.
Ireland	Vocational training opportunities scheme (VTOS)	WRC Social and economic consultants Ltd., 1994	This is a detailed evaluation of a nationwide programme of training for the long-term unemployed that contains good impact analysis. The evaluation is integrated with the programme of reform.
Italy	<i>Il nuovo apprendistato e la formazione</i>	Angeli, 1999 ISFOL, 2000	This reform is system-wide and is representative of renewed interest in apprenticeship across several countries. The evaluation focuses on the infrastructure of the reform. Other papers outlining system-wide reforms provide perspective on the apprenticeship and some impact analysis.
Netherlands	Adult education and vocational education act (WEB)	Borghans and Heijke, 2002; De Bruijn et al., 2001 Nieuwenhuis and Shapiro, 2003	This reform is included as it is a case study in the third research report. The reform is nationwide and involved decentralisation of management changes to VET programmes, funding changes. The evaluation has a formative function. Several as yet unpublished papers have been made available.
Portugal	<i>Sistema das escolas profissionais</i>	Montavavao e Silva et al., 1996	This is a reform of IVT. The diverse range of vocational training is to come under a more coherent approach but includes the introduction of specialist schools. The new schools are autonomous although there is some regulation of them. The evaluation is phased and commissioned from an international source.

Portugal	<i>Despacho Conjunto – do Ministério da Educação e para a Qualificação e o Emprego – No 123/97 de 16.6.1997</i>	Ministério da Educação e para a Qualificação e o Emprego, 1997	An annual ministry-led evaluation programme that involves a 'bottom up' approach is included. Finance features in this evaluation.
Spain	<i>Memoria Banance – I Acuerdo Nacional de Formacion Continua 1993-1996</i>	Forcem, 1999	Developing a culture of professional development is the goal and evaluation of impact of the first three years of promotional activity is provided. The reform is wide ranging and involves many stakeholders.
Sweden	<i>Reformeringen av gymnasieskolan</i>	National Agency for Education, 2000 Ekstrom, 2002	A system wide reform of upper secondary education (qualified vocational education, QVE) with a strong focus on IVT. A full-scale evaluation of this reform has been written and is currently with the sponsoring organisations. It was not available at the time of writing.
Sweden	Swedish youth labour market programmes	Larsson, 2000	Two programmes involving training are compared. The evaluation has an econometric design to determine the effects of these programmes on earnings and reemployment probability.
UK	Work related further education programmes	Metcalf and Dawson, 1995	A study of reforming planning and financing of IVT and CVT in colleges. The evaluation uses an interesting comparative method which is quasi experimental. Impact measures are included.
UK	The Union learning fund	Cutter et al., 2000 Antill et al., 2001	This is an evaluation of an initiative that includes social partner involvement, VET capacity building and financing. It deals with basic skills training in particular. The evaluation is a blend of survey and case study and runs in three annual phases thereby offering progressive reporting. Many output measures discussed.
UK	Technical and vocational education initiative	Department of Employment, 1994	This is a large-scale curriculum initiative to reorientate learning to more active approaches and to increase the profile of vocational education. It has been evaluated extensively.
UK	Modern apprenticeships	Economic Research Services Ltd, 1998; Coleman and Williams, 1998	A new apprenticeship scheme that has been evaluated for impact and over a period of three years. Learners and employers are surveyed in detail.
UK	Funding systems and their impact on skills	Felstead and Unwin, 1999	This paper has no VET reform at its heart but paves the way for changing the way VET is funded. Two funding systems for IVT are compared and evaluated.

3. Criteria for judging reforms

One of the challenges in developing a meta-study of evaluation reports is to set analytical criteria to make the study systematic. A criterion-based review makes it possible for readers to judge the value of the review for their purposes. If necessary, readers can extend the review by refining the analytical criteria and altering judgements of reforms accordingly. Developing criteria for this study began in preparing the proposal to carry out this work. The proposal made the point that it is simplistic to use criteria based on the conventional anatomy of reforms (conception, consultation, initiation, pilot phase, implementation, review, refine, etc.). Instead it was proposed to look more closely at the evaluation procedure as well as the reform and to focus strongly on impact. For the reform this meant looking at the aims and objectives of it closely and, for the evaluation, looking for baseline assessment, comparative approaches and impact or performance criteria. It also meant looking for unexpected outcomes and system-wide implications. A set of criteria was developed using a range of literature (e.g., Barrett, 1998; European Commission, 1999; Plewis, 2001; World Bank, 2002) and expert sources within the group of researchers working with Cedefop on the third research report and within the Qualifications and Curriculum Authority (QCA).

The second stage was to seek the views of other experts of the draft criteria. Amendments were made accordingly and a working document produced. The structure of this set of criteria was based on the chronology of an evaluation process – before the reform was implemented,

during implementation and post reform. To these three a set of general criteria were added on the evaluation process itself. Later, it was considered advantageous to introduce a further set of organising headings to make it clear that a chronological approach was only one way of organising the criteria. This additional set of headings is as follows:

- (a) context: what external influences are likely to influence?
- (b) goals: how are the aims, objectives, milestones described?
- (c) planning: what is discussed in anticipation of what might happen?
- (d) management: what is done to make the programme function effectively?
- (e) drivers: who is shaping the evaluation?
- (f) impact: what is the effect of the evaluation?

When these headings are applied and the 40 criteria within the chronological headings are allocated the coverage is as follows: context (6), goals (5), planning (4), management (13), drivers (6) and impact (6). This distribution reflects the focus of this study with the emphasis on drivers and impact being served by a higher proportion of criteria than would normally be expected. The application of these criteria to the reports of reforms is discussed later in this report.

In the next chapter the emerging observations, generalisations and issues are discussed. The headings used to group the common evaluation criteria (see annex) are used to organise the discussion.

4. Evaluating VET reform

The selected reforms and their evaluative reports represent a large body of evidence on the recent history of change in VET in Europe. In this chapter some review of VET reform is presented followed by a synopsis of findings on changes to the financing of VET. A general discussion on evaluating the effects of VET reform concludes this chapter.

4.1. VET reform

The reforms studied exemplify a wide range of VET activity. They cover vocational preparation in schools, full-scale initial vocational training (IVT) programmes, continuing vocational training (CVT) in work, CVT for those who are seeking employment, retraining activities, short programmes, long programmes and qualifications from the International Standard Classification of Education (ISCED) levels 1 to 5. The reforms aim to change the management of VET, financing of training, the learning environment for VET, the teaching approach, the content of courses and, of course, to boost outcomes of training. The image of VET gained from studying the reports is one of tremendous diversity, with multiple purposes closely related to the dynamics in families, communities and institutions. The breadth of VET means many people are agents in supporting, designing, delivering or assessing programmes and these people are generally well consulted in reform development. They are also generally well consulted in evaluation programmes.

In Cedefop's second research report (Descy and Tessaring, 2001) it was established that there was a general move away from State-led VET and market-led VET to new forms of collaborative steering arrangements, possibly through professional associations and networks. There is considerable evidence in the sample of reforms used in this study that such a trend is continuing: examples include designing new semi-autonomous institutions for IVT in Portugal and policies to decentralise provision of State-funded VET in the Netherlands and the Nordic countries.

The institutions that distribute State funds to meet the needs of individuals and employers seem to be the key brokers in the new professional associations and networks.

All reports studied described the aims of reforms; these aims reflect the wide range of VET applications discussed above. For this research study, it is important to look at the relationship between the aims of the reform and the objectives of the evaluation programme associated with it. This is key for impact analysis and is discussed in detail later in this chapter.

Almost all the reports studied examined changes to the way VET is organised. Examples include new types of IVT programmes (chiefly apprenticeship schemes and schemes for making general education more vocational), CVT programmes (chiefly State-funded schemes and other ways of organising funding). Additionally attention is being paid to reorganisation to improve social welfare and the autonomy of the individual. Innovation in VET provision is evident and takes many forms, for example in the first year of developing the Union learning fund in the UK the project had innovative ways of delivering VET at its core.

The reforms chosen for analysis were primarily selected as examples of impact studies and this will have skewed the representative nature of the reforms as depicting types of VET activity. However the reforms do reflect the span of national programmes and it is notable the focus on reforming IVT is strongly represented in these reports. Usually this involves developing IVT alongside general education and allowing those with a disposition for more applied study (experience of work, content of courses, teaching, learning and assessment styles) to opt into vocational programmes. By doing so it is hoped these people will be more motivated to learn and will develop useful skills sought on the labour market. Consequently, they will be less likely to drop out of school and lose the advantages of a good initial general education. Some believe these programmes result in a diminution of the volume of general education taught to the participating

cohort. There is also evidence of some permeation of teaching styles and resources from IVT provision to general provision when these programmes are taught in the same institution.

Access to higher education also features in the selected reforms. Returning to the changes in upper secondary IVT programmes, a common objective is that students following these new programmes should have access to higher education, so more time is needed to develop the vocational provision while retaining a core of general education. Several reforms include additional time for study in upper secondary education – typically another year. This additional time is needed for learning from workplace experience and for developing and assessing projects or assignments. In some reforms studied legal changes provide a right for individuals graduating from such programmes to have a place in a higher education institution. One reform that aims to make higher education provision more desirable and accessible to students who retain this interest in vocational programmes.

Some reforms aimed to create new learning centres, others aimed to change participation and pedagogy in existing centres. Sometimes the reforms to content, pedagogy and VET organisation are sufficiently extensive to create a new learning environment based in an existing institution. Thus it is sometimes difficult to isolate reforms for creating new environments from those simply changing content or pedagogy. In this study new learning environments were created by:

- (a) the Austrian reforms where students are educated in new *Fachhochschule* where expertise could be concentrated on vocational courses;
- (b) introducing apprenticeship programmes in Italy and the UK where employers and training providers established new ways of blending general learning with work-based learning;
- (c) the reforms in Portugal where new semi-independent regionally controlled schools for IVT were established alongside existing provision;
- (d) trade unions in the UK where they negotiated

new provision with a different rationale and management.

Many reforms aimed to create a new purpose for learning programmes by developing existing institutions. This could be by altering management structure, financing or course provision. Examples of these reforms include:

- (a) regionalisation of management of institutions delivering upper secondary education in Denmark, the Netherlands, Finland and Sweden;
- (b) improvement in CVT in former East Germany to meet better the needs of workers and employers;
- (c) increasing emphasis on work experience in Denmark, Ireland, the Netherlands, Finland and Sweden.

Several reforms aimed to change the content and structure of programmes within existing institutions that were not required to change significantly. Examples include:

- (a) experimenting with new content in the Italian apprenticeship programme;
- (b) the UK reform of the general curriculum in schools to create a more vocational focus;
- (c) inclusion measures in the Irish vocational training opportunities scheme (VTOS);
- (d) introducing new forms of assessment in Finland and Sweden;
- (e) developing new schools leaving qualifications in Ireland.

We can classify these reforms in learning and learning environments as shown in Table 2.

Many reforms allocate time to general education and sociocultural studies as well as development of specific technical skills – the balance is sometimes contested. For example, in the Portuguese reform employers and parents see teaching and learning about general social issues reducing the time that could be usefully spent on work-based assignments and skills learning. The designers of programmes containing an element of general education clearly have to strike a balance between the expectation of users and the strong evidence supporting inclusion of elements of general education in IVT provision.

Table 2: **Classification of reforms to learning and learning environments**

Change	Aims of change	Examples
Create additional new institutions	Concentrate expertise, diversify VET provision, attract students, meet skills need.	• Austria (Wimmer, 2000) • Portugal (Montavavao e Silva et al., 1996; Ministério da Educação e para a Qualificação e o Emprego, 1997)
Blend existing learning environments into a new kind of provision	Produce more effective programmes, shift control of training to employers, tailor courses to meet skills needs.	• Ireland (WRC Social and economic consultants Ltd., 1994) • UK (Department of Employment, 1994; Metcalf and Dawson, 1995), Spain (Forcem, 1999) • Portugal (Montavavao e Silva et al., 1996; Ministério da Educação e para a Qualificação e o Emprego, 1997)
Engage social partners in planning and management of provision	Innovate in programme design (focus on learner), coordinate range of existing provision.	• UK (Cutter et al., 2000; Antill et al., 2001) • Spain (Forcem, 1999) • Portugal (Montavavao e Silva et al., 1996; Ministério da Educação e para a Qualificação e o Emprego, 1997)
Develop the school curriculum and assessment and qualifications	To make curriculum relevant, assessment authentic, recognise informal learning, motivate learners, bridge initial education and work situations.	• Sweden (National Agency for Education, 2000; Ekstrom, 2002) • Finland (Finnish Ministry of Education, 2000) • the Netherlands (Borghans and Heijke, 2002; De Bruijn et al., 2001) • Denmark (Shapiro et al., 2001; Kaasbøl et al, 2001; Nielsen, 1997) • Ireland (Irish Ministry of Education, 2002)
Improvements in the content and teaching of specific programmes in an existing type of institution	Work force development, drive up productivity through deployment of new skills, improve progression opportunities to work and in work.	• Denmark (Shapiro et al., 2001; Kaasbøl et al., 2001; Nielsen, 1997) • UK (Department of Employment, 1994)
Develop teaching approaches to motivate and retain learners	Reduce exclusion from work and costs of benefits.	• Ireland (WRC Social and economic consultants Ltd., 1994)

4.2. Evaluation

Evaluation reports generally describe the subject of the evaluation, outline objectives, briefly describe the method and then report findings and suggest implications. It is usually not possible to find details of the evaluation process, the practical constraints, the compromises, the adjustments, the pressures for results, the consideration of different ways to interpret results and how results and conclusions are protected from misin-

terpretation. Reading reports will not reveal this information. Academic research into evaluation methods usually contains discussion of the broader issues but this discussion will normally be out of the context of a specific reform. It is possible that only through discussion with planners, evaluators and programme managers will we learn about the evaluation process itself. Therefore, in reports such as this there is a gap between full knowledge of the reform, the interventions, the evaluations and what is reported in literature.

Evidence from evaluations of VET can be useful for:

- (a) inform government decisions (including allocating funds);
- (b) improving employer decisions on training;
- (c) informing individuals of their options;
- (d) designing new programmes;
- (e) refining programme design;
- (f) improving programme targeting;
- (g) identifying ineffective programmes;
- (h) encouraging public debate about VET.

Stern (2003) summarises the purposes of VET evaluation under four categories:

- (a) accountability for policy-makers;
- (b) development for programme improvement;
- (c) knowledge production and explanation;
- (d) social improvement and change.

Impact evaluation forms part of only the first of Stern's categories where he looks for 'impacts, outcomes, achievement of targets and value for money'.

When a VET programme is created or changed it is important to know if the new programme works. Programme evaluations therefore require (usually quantitative) measures of impact – to judge if an intervention has worked or not – and (usually qualitative) information to discover the reasons why the intervention has succeeded or failed. Both these measures are important but in this study the first measure is of prime interest. Interventions based on reform of VET can rarely, if ever, be isolated from other contextual influences and this makes measuring impact highly complex. To help focus hard on impact there are two questions that might be asked:

- (a) what are the changes in VET (including changes to those involved in VET and the contexts in which VET learning is applied) that have occurred during the period of intervention and after the intervention?
- (b) is there evidence to link these changes directly or indirectly to the VET intervention?

Restricting this study to reports that provide answers to both these questions eliminate reports of process evaluations that analyse how a programme operates and focus on impact evaluations and causality in particular. The common evaluation criteria (see annex) help with this sharp focus on the two questions.

Literature shows the way evaluation is considered varies greatly. In fact the range of practice is

so wide that it is difficult to understand how a single term like 'evaluation' can cover all instances. There is a need to distinguish research according to purpose, i.e. accountability, efficiency, functioning (impact) and valorisation or as Stern (2003) puts it: accountability, programme improvement, knowledge production and social improvement.

One striking difference in evaluation programmes is the distinction between those that aim to answer the question 'has it worked?' rather than answer a series of questions based on 'is it working?' In other words, end-loaded summative evaluation relative to integrated formative evaluation. Evaluation helped the reform process along in just over half of the cases sampled in this study. On the potential of the evaluation to yield information for the reform process, the reports used are once again evenly split between those that have been designed to inject a sense of progress or impact at any stage during the reform implementation and those that have not.

Linked to the question of timely intervention is the question of deciding when it is reasonable to make a measurement of effect after the reform has become established. It is rare to find an evaluation that has measured changes to participants in VET programmes a year (or longer) after completing a VET programme. One of the best examples is the Swedish reform of IVT in upper secondary education (Ekstrom, 2002). In the three academic years between 1988 and 1991 the current Swedish upper secondary vocational pathway (three years duration) was piloted with students occupying 11 200 study places. The career development of these students was tracked by measuring the time spent in upper secondary education, level of university enrolment and the rate of educational or employment inactivity. However, a large proportion of Swedish youth were not eligible for this pilot programme as they were following the usual two-year vocational programme. Consequently, 'natural experiment' was created where important influences other than education could be assumed to be the same for both groups of students. Attempts were also made to balance various regional differences.

Results showed that the students in pilot areas, where both two and three-year programmes were available, were more likely to spend three years in upper secondary education. Higher education

enrolments were higher for the pilot group with three years of vocational upper secondary education. The effect of a third year of vocational education on the rate of inactivity of individuals was not significant. One possible explanation is the Swedish labour market declined during the pilot phase and students on two-year programmes had time to establish themselves in jobs before the decline took place. The report of this evaluation highlights the effects of changes in context over a period of years following an intervention.

The low number of evaluations such as the Swedish study that go significantly beyond intervention has major implications for this review of

impact studies since the subject or focal point of any effect changes with time. This effect is discussed later. Many reports suggest that an evaluation of the long-term effects of reforms is useful but only six reports attempt gathering data after intervention is complete and analyse and report on them. However, several evaluations, possibly five more are in progress and will continue to measure progress and/or impact.

Looking across a spectrum of VET reforms and evaluation methods it is useful to look at the type of reform, the kind of desired outcome or effect, the evaluation style and the research tools used. This overview is described in Table 3.

Table 3: **Broad review of VET reforms and evaluation methods in this study**

Type of VET reform	Intermediate impact	End point impact	Research tools used	Evaluation style
Development of CVT D, EL, E, NL	Increased participation of adults	Higher skills levels, economic growth	Questionnaire survey of institutions and individuals, satisfaction survey, case study exemplification	Continuous monitoring and reporting programme
Reform of post compulsory IVT DK, NL, FIN, S, UK	Content development, pedagogic development, institutional development, regionalisation	Increased participation, good progression for graduates	Questionnaire surveys, consultations with stakeholders	Monitoring and formative evaluation
New qualification development IRL, I, UK	Increased participation	Close skills gap, higher skills levels, economic growth	Questionnaire survey of institutions and individuals. Satisfaction survey	Monitoring and formative evaluation
Cohort review F	Identify successes and problems	Modify provision	Periodic survey	Independent monitoring
VET funding review UK	Changes to funding distribution system	More efficient and responsive system	Literature review, interview, focus groups	Single event evaluation
VET institutional change A, P, FIN, S, UK	Establish popular new learning institutions, make institutions more responsive	Capacity building, improve skills supply	Stakeholder consultation, case study exemplification, satisfaction survey	Formative evaluation
Inclusion measures D, IRL, S	High take up of programmes	Lower long-term unemployment	Stakeholder consultation including use of controls. Post programme review	Mix of evaluation styles

The type of evaluation tool does not necessarily link with the type of reform but in this study it is clear some types of VET reform, for example VET linked to inclusion measures draws on a broader range of evaluation tools than many others. Almost all reforms use some form of user response questionnaire, usually postal but sometimes telephone based, to measure effects of programmes. These usually have three main parts:

- (a) personal and demographic background information of the subject;
- (b) ratings by subjects of aspects of the VET provision they are using or have used;
- (c) judgements by subjects about future prospects for employment or learning progression.

Extending these questionnaire-based approaches so analysis can have a greater impact focus is discussed later in this report.

In some evaluations, there was an issue of timing of feedback from the evaluation programme to the developing reform. This is restricted to evaluation programmes where evaluators are not involved in the programme itself. Typically, data will have been gathered, analysed and then reported but in the meantime the reform will have moved on and the data appear to be, at least in part, redundant. A good example in this study is the evaluation of funding of work-related college programmes in the UK. The ways in which colleges involved in the reform produced business plans for future provision soon spread to other colleges not involved in the study. By the time the evaluators reported, the ideas permeating between colleges had become a flood and the report was more documentation of progress than something that might shape dissemination of outcomes to other colleges not involved in the study.

Having reviewed generally the nature of VET reforms and the ways they have been evaluated this report now begins using the common evaluation criteria to examine the evidence base in more detail. First, a review of criteria is described followed by deep consideration of measuring the impact of VET.

4.3. Context of reform

Almost all reforms were set in the context of broad changes brought about by governments. Whether this link to national policy was a trigger

for the reform or whether in reporting on evaluation the authors attempted to locate, post hoc, the reform in a wider setting is usually not clear. However, most reforms are fundamental and it is difficult to perceive them as developing independently of central government policy, for example the Nordic reforms of upper secondary education.

Many reports describe the forces promoting change in VET. These include:

- (a) the political wish or mandate to reorganise VET;
- (b) economic necessity;
- (c) relief of social issues;
- (d) the need to find more equitable ways of funding VET;
- (e) the need to overcome bureaucratic barriers;
- (f) feedback from enterprises about skills needs and shortages.

Legal requirements for new structures to be established and for new individual rights to training undoubtedly carry forward VET reform but it is difficult to establish whether legal requirements are a driving force. There are numerous examples of reforms driven by law (prime examples come from Spain, Italy, the Netherlands, Finland and Sweden). Research on the driving forces for VET reform is worthy of further exploration on a European basis. There are insufficient data in literature on how different agencies drive VET reform in the reports used in this study and it may be more fruitful to gather information on drivers of reform independently and outside impact studies.

4.4. Goals

All reports included some information about goals or aims of the reforms being evaluated. These aims were usually stated in broad language and not related to impact on individuals or organisations. Typically, aims referred to improving economic performance by raising skills levels or improving access to the labour market for disadvantaged groups or developing a culture of professional development. Separate identifiable goals for the evaluation process were also present in half the reports but these were often worded to evaluate the process of reform rather than the outcome of reforms. For example, to report on the future direction of the reform, to

improve reform management or resource allocation. About one third of studies were clearly defined for measuring outcomes with stated aims of impact on specific stakeholders.

One strong differentiator of reports is the extent to which the goals of reforms and evaluations referenced theory-based reasoning for taking the proposed course of action. One third of reforms argued for a new form of VET by using evidence from research, another third made no such case and for the remaining third it is difficult to determine where the arguments for new forms of VET originated. A theory-based reform typically has theoretical proposals such as 'knowledge is situated in specific contexts' and include mechanisms for translating theoretical knowledge into outcomes such as 'use a range of contexts for learning ideas' and the outcome could be 'greater likelihood of employment' or 'better performance at work'. Griffith (2001) provides useful information on theoretical perspectives in evaluation studies.

It was not possible to make judgements on how aims changed during evaluation. One study made it clear the direction of the reform had changed significantly during its implementation and the purpose of evaluation had also changed. As stated earlier only a few reports described effects of interventions some time after the intervention had taken place.

4.5. Planning of reforms

Establishing a baseline for analysis is a basic requirement for impact studies. Most reports contain information and justification of a post hoc rationale for accepting a baseline measure. It was not possible to judge whether any projects had established the baseline data to be used to judge effect before a reform got underway. There was a paucity of reports describing establishment of an ex ante evaluation process.

There were indications in about half the reports that the evaluation process was designed to aid development of the reform and was to take place simultaneously with the reform.

Reports usually set the reform in a wider political or economic context but it was not possible from the reports alone to deduce there was a sense of coordination and coherence within a set

of reforms in one country. However, many reforms centred on institutional change which depended on changes in governance and funding of institutions – a coordinated approach was a prerequisite in these cases.

Dissemination of findings from evaluation processes was rarely discussed. This might signal such processes were not defined at the outset of the project.

4.6. Management of reforms

Almost all reports described the structure of the reform process but very few described in any detail how the reform was managed. Those that included descriptions of management structures and responsibilities used them to describe the reform itself. For example, in the Portuguese study the setting up of new institutions required new regional management structures. These were seen in the evaluation as the subject of the reform and not as a means of managing the reform. There was little detail on management infrastructure, steering groups and job specifications for key workers. More information is provided on accountable organisations or committees: a ministerial report on the Portuguese reform offers some insight. Annual evaluation (effectively a ministerial meta-study) is based on a hierarchy of active committees. At local level, a board of teachers evaluate performance as they see it. These reports are forwarded to a regional board that analyses the teachers' reports and prepares a regional evaluation. This is then passed to a national council that summarises regional reports and presents an annual progress report to government.

Common evaluation criteria contain reference to the commitment of donor organisations. In fact few reforms were independent of government and there was almost always an unstated assumption that the reform would be supported to completion, that is, until a new structure was in place. There were some exceptions, for example in the UK Union learning fund the provider was the UK government and over a three-year period the efficacy of the scheme was to be established. After the first year negative results would have led to abandoning the scheme. In developing IVT schools in Portugal the first phase was to estab-

lish whether it was worth developing the programme further. It was never evident from reading the reports that the donor influence was affecting measurement of effectiveness of a reform programme. However, some researchers believe such influences exist and they can be applied when determining the scope and style of evaluation and in reporting findings (Martin, 2001).

Risk management of reforms was invisible in evaluation reports. This is perhaps not surprising since the evaluation programme itself was possibly part of a risk management strategy. Even with government-led reforms where success is highly desirable for political reasons risk considerations had a low profile. It is possible in the quest to obtain reports concentrating on impact evaluation, that good reports assessing process risks were ignored. Whatever the reason and for management appraisal in general, there are signs that programmes are finding it difficult to establish this aspect of evaluation. One reason may be difficulty in judging management in such large reforms. In each country there may be few experts with experience of such reforms.

Only one quarter of reports described a pilot phase for the reform. However, many reforms (e.g. VTOS and leaving certificate applied in Ireland, modern apprenticeship in UK, and the development of Swedish vocational programmes) were voluntary for users and therefore developed over time starting with a relatively small base before expanding to full scale. There was also little evidence of evaluation of the means by which reforms were expanded from pilot to full scale or disseminated more widely.

In many cases, how other reforms (e.g. to benefit systems, guidance services, regional management) interfaced and supported the reform were described. Many of these descrip-

tions arise because measuring effect often demands an explanation that goes beyond the effect of the reform itself. For example, the Greek evaluation of training programmes explains that those undergoing training have a better chance of gaining long-term employment. These people often use more sophisticated job application techniques than those who do not undergo training and they find they can access other means of government support.

4.7. Drivers

As part of the contextualisation of a reform the key stakeholders in developing, carrying out and benefiting from the reform were often identified. The roles of different groups in making a project a success are almost always described. However, interaction between these stakeholders and the changes in their relative authority within VET was never discussed. This kind of discussion rarely finds its way into evaluative reports. Such changes in the power to influence are crucial for the long-term sustainability of VET reforms and consequently their impacts on stakeholders. Analysis of changes in the structure and dynamics of managing VET change is more likely to be included in process evaluation reports so these have been intentionally excluded from the set selected for this study.

4.8. Impact

Focus of the research on impact studies, the ways in which effects are measured and the problems faced are discussed in full in the next chapter.

5. Financing VET: reforms and evaluation

Keeping the financing of VET under review is important. Individuals, society and the economy benefit from an effective VET system and the State, enterprises and individuals have an interest in investing in VET to gain rewards later. The multiple funding sources for VET and the range of programmes requiring financial support make this field highly strategic – reallocation of resources from one programme to another is usually a strong indicator of values and priorities held by funding bodies. It can simply mean one programme is more effective or more important than another; it can also mean policy has changed, priorities are now different or the demand side of training is changing. Thus, to evaluate shifts in financing is a major contribution to evaluating VET itself. It is interesting to note, however, evaluation of the effects of funding of VET reform was discussed in any detail in only 5 of the 19 reports selected. One report was specifically about funding and two more were on judging whether a reform had yielded value for money. The paucity of empirical studies of the efficacy of financing procedures has been reported (Kath, 1999). Of particular note is the shortage of systematic longitudinal studies. There are major methodological difficulties in establishing such studies (Barrett, 1998). First measures of impact have to be developed and agreed as valid. Then the methods must be capable of isolating these effects from other influences and the costs (direct and indirect) and expenditure (often from multiple sources) have to be characterised clearly and unambiguously. Kath (1999) suggests that no cross-national comparison of the effectiveness of VET financing systems is yet possible because of these difficulties.

5.1. Financing reforms

There were some interesting examples of using a funding review to improve the targeting and effectiveness of VET (for example the work related further education review in the UK and the change in financing arrangements for IVT in

Portugal and Sweden). In the UK example, the reform required a sample of colleges to change the way they planned their future provision, by consulting employers more thoroughly, establishing partnerships in planning and evaluating effectiveness collaboratively. This way of determining funding needs has now become part of the UK system of funding VET.

Another funding reform signalling reappraisal of the way central funds for VET are distributed was providing incentives for additional (levered) funds from independent providers. The example below shows how this expectation was measured and reported as an effect of the reform.

Some reforms include major changes in the way VET is funded; most of these concern regionalisation of control and devolution of funding responsibilities. Others involve using private funding in public provision and yet others have introduced new ways for individuals to pay more or less for their own training. The upper secondary reform in Finland has a specific focus on financing. In the new Finnish system, collectives are now required to distribute funds for training – these funds were previously held by institutions. Unfortunately, evaluating the effectiveness of this change in funding arrangements is limited to a user view of how the new systems have worked. This has revealed, for example, that users have found it too complex. Sustainability of funding systems is discussed in several reports.

Some innovative funding mechanisms are reported, one concerned the UK national government funding a union-led initiative to explore ways local networks of workers and trainers could develop systems for training for local needs. Essentially this was a union training capacity building exercise that fits well with life-long learning objectives. The scheme expanded over the three years of evaluation and was successful in creating a range of training provision not previously available and as reported above, in leveraging funding from additional sources. The scheme was also incorporated into UK mainstream provision.

Table 4: **Measuring the impact of new funding arrangements**

Year of programme	Year 1	Year 2	Year 3
Number of projects with levered funds	16	13	30
Proportion of total projects	36 %	20 %	32 %
Average amount (per project) levered GBP 1000s	28	43	54
Source: Union learning fund (UK).			

5.2. Evaluation of financing

Using evaluation data for resource allocation is an issue for governments and large-scale funders of VET (Summa, 2000). Using evaluation data as justification for changing resource allocations is difficult as evaluations tend to recommend changes to single programmes and do not provide evidence of the effectiveness of VET across different contexts. Many reports simply state costs of programmes which is of little evaluative use, but sometimes data are given as trends over the period the reform operated. This is sometimes accompanied by a description (rather than an evaluation) of the funding distribution system (e.g. VTOS in Ireland). It is possible that absence of financial data shows that analysis leading to evaluating the value-for-money of reforms is problematic. This is likely to be particularly difficult when the gains of the reform are judged against the system that existed before the reform was introduced. The key issues are likely to be absence of controllable variables and the consequential lack of a baseline for measurement. The Irish VTOS report attempts to show the differences in scale of funding for VTOS as opposed to other VET for the long-term unemployed and discusses the numbers of participants 'processed' by the different schemes. This comparative work is helpful but does not give a clear indication of value-for-money.

The econometric studies included in the sample of reports simplify judgement on whether one reform was more effective than another (within the bounds of how much it is possible to know of what effectiveness means) but rarely go further and comment on the funding process itself or of value-for-money of VET provision outside the comparative study.

Some reports include information on how learner finance issues affect take up of VET provision. The Portuguese report outlines how student financing is affecting retention. A law governs how student fees are set and the extent to which a school can borrow money and sell products. Evaluation showed the financing model to be faulty as it was too short term, highly bureaucratic and heavily dependent on government subsidy. The report also highlighted serious weaknesses in evaluation methods used to judge the effectiveness of the finance process. The Finnish report also details some research into financing measures. New upper secondary vocational provision is funded 60 % by the State and 40 % by regional authorities. The report also explains the possible commercial funding of provision and its limitations, although once again there is no evaluation detail.

The evidence in these reports on financial aspects of reforms is limited. There is some econometric analysis with a scientific basis which is probably more reliable for judging the effects of VET financing.

6. Levels of evaluation

Measurement of impact is the key focus of this report. It may be important to distinguish between impact and outcomes. Literature suggests the most accepted classification of outcomes of training is the four-level system first described by Kirkpatrick (1998). These are:

- (a) reaction to training: this determines how much the learner might learn;
- (b) learning: a measure of how much is learned and therefore available to transfer to the job;
- (c) behaviour transfer: how much work behaviour changes as a result of training;
- (d) organisational results: in what ways and how much the organisation benefits from changed behaviour.

These four sets of outcomes are hierarchical and maximising each one provides a better base for the next. Grubb and Ryan (1999) take a broader view; they describe outcomes using the following categories:

- (a) for the individual, e.g. economic, wages, employment pattern, progression, developing learning and skills;
- (b) outcomes of other agents, e.g. productivity;
- (c) welfare and social programmes: reduced dependence;
- (d) social outcomes, e.g. reduced criminal behaviour.

A way of reconciling these two approaches is to consider all that is planned and delivered as VET as 'intermediate' outcomes, for example, performance of the learning infrastructure, efficiency and equity of learning providers and qualification target realisation. As stated earlier, programme evaluation, as opposed to impact evaluation, will start from changes to infrastructure as an outcome and then progress to changes in institutions and programmes and finally to impact on individuals. In fact, according to Plewis (2001), the final destination of impact is with the individual receiving VET, their families, their community, organisations and society generally. It is this model of assessing impact which underpins the analysis presented in this report although it is clear that as one moves

away from the individual other variables and contextual factors often mask the effect of the specific VET intervention.

6.1. Establishing a baseline

There are no absolute positions for measuring the impact of VET intervention. Established systems – no matter how they underperform – have an impact in their current form. Around half the reports described the process of establishing a control group or making some measurements before intervention began. Some reports, not surprisingly those using econometric techniques, made strenuous efforts to make a convincing case for having measured an absolute change in a condition as a result of the intervention. There is some evidence that large-scale reforms do not appear to make ex ante evaluations or do not conduct them in sufficient detail to make impact evaluations later ⁽²⁾.

As stated earlier every report made an explicit statement of the aims of the evaluation in relation to the reform. Most describe their objectives in fairly general terms, for example:

- (a) to support changing requirements for training (D);
- (b) to develop earnings and career prospects (D);
- (c) to promote innovative activity in training (UK);
- (d) to maximise the value of study for employers and individuals (UK);
- (e) to improve the supply of intermediate skills (UK);
- (f) to diversify VET provision (P);
- (g) to make the curriculum more flexible (P);
- (h) to experiment with the content of syllabuses (I);
- (i) training to help you get the job you want and keep it (EL);
- (j) to develop student autonomy (S, NL);
- (k) to lower entrance barriers to education (FIN);
- (l) to create a culture of professional development (E).

Other reforms had more specific objectives:

- (a) to improve student results (UK);

⁽²⁾ See, for example, the final assessment report for assistance funded under Phare VET reform programme in the Czech Republic, OMAS Consortium Middle Region, 1999

- (b) to reduce long-term unemployment figures (IRL);
- (c) to retain more students in post compulsory schooling (IRL);
- (d) to reduce youth unemployment (S).

To measure impact it is probably helpful to transform this type of aim or goal into more specific objectives. This often takes place when the objectives for the evaluation programme are drafted. Some examples from evaluation reports demonstrate this:

- (a) assess the probability of getting a job, the rise in earnings, the probability of higher-level study ... all relative to a control group (S);
- (b) how well VET provision meets individual demand (NL);
- (c) inspect evidence of practice in schools measured against common criteria (IRL);
- (d) educational gains in qualifications (IRL);
- (e) comparing special VET provision and normal VET provision for satisfaction ratings (UK).

This area of evaluation design – definition of evaluation objectives – is where an evaluation programme can be optimised for generating impact.

6.2. Impact and timing of measurement

Impact can usefully be considered in two discrete dimensions. The first is the subject of impact (e.g. individuals, institutions), the second is a temporal dimension.

If we take, for example, an employer-led training programme as the VET intervention we may consider the subject of impact as any of the following:

- (a) level of competence of an individual worker;
- (b) capacity of a work group to work differently (flexibly, creatively);
- (c) capacity of the individual or work group to disseminate new knowledge;
- (d) productivity of a work unit;

- (e) stability of staffing;
- (f) manageability of a work unit;
- (g) profitability of the company.

There will be more subjects of impact for different specific situations. The effect of the intervention also changes with time. The change can be continuous, for example the content of the VET intervention can dissipate or the skills learned can lead to learning new skills. The change can also be categoric in nature and be seen in distinct stages. For example, the effect on an individual could be focused on their reaction to learning during a training programme or about getting a job based on the course in the weeks following a programme or about promotion and further learning in the longer term.

The impact evaluations included in this study generally concern themselves with the immediate effects of a programme on VET provision or customer satisfaction with their VET programme. They generally have a weak time dimension; they are timed to measure effects during the intervention or soon afterwards. A Swedish evaluation extending several years after intervention has already been outlined above.

VET reform might change the situation of an individual (e.g. employability, earnings, promotion), the VET structures and systems (e.g. developing trainers, financing, qualifications) and associated supporting arrangements (e.g. guidance, skills needs analysis). This kind of analysis is common in evaluation literature (e.g. European Commission, 1999) and is a useful way of considering the reports included in this study.

6.3. The individual

The individual is the focus of many reforms and the views of individuals are sought in most of the evaluation programmes included in this study. However these views are often sought on how the individual feels about the intervention, for example:

The level of satisfaction from vocational training programmes was found to be quite high. Out of all trained respondents, 90.6 % were very/quite a lot satisfied by the infrastructure of training centres, 89.3 % by the quality of training, 89.8 % by the trainers' standards and 83.8 % by the usefulness of the training provided. Overall, satisfied respondents reached more than 90 % (91.8 %) and remained high level among all demographic groups.

Source: Greek report of VET effectiveness.

While this information is important and useful there may be dominant short-term influences operating that, in time, may decay. Some evaluations attempt to draw out a longer-term

perspective from participants and seek information about how individuals feel relative to others who had not participated in the intervention. For example:

Table 5: **Proportion of *Fachhochschule* leavers who feel disadvantaged by their counterparts in 'normal universities' with regard to selected aspects**

(%)

Income	74.4
Prestige	58.4
Career chances	38.2
Opportunity for progression in company	24.4
Source: Austrian reform of higher education	

Taking this further, the effect of an intervention might best be measured by difference in opinion between an individual participating in the

programme being reformed and one that is not. This might yield more reliable information about the relative effect of the intervention. For example:

Table 6: **How much an apprenticeship course is helping to...**

(%)

	Intervention sample	Comparison sample
get skills and qualifications	77	72
find out about working life	63	10
gain entry to chosen career	59	57
decide on a career	52	38
find out how firms work	48	20
find out about job opportunities	27	10
Source: UK modern apprenticeship study		

Both the differences in response rate and the similarities can be useful. Moving on from views about the future, and possibly increasing the usefulness of

data on effects, another style of impact evaluation looks at actual outcomes as opposed to views and opinion about future outcomes. For example:

Table 7: **Increases in salary from the first employment in 1992 to employment in 1997**

(FRF)

Qualification	Salary 1992	Salary 1997	Difference
<i>Bac technologique industriel</i>	5 500	7 500	2 000
<i>Bac professionnel industriel</i>	5 500	7 300	1 800
<i>Bac général</i>	5 000	6 500	1 500
<i>Bac technologique tertiaire</i>	5 200	6 000	800
<i>Bac professionnel tertiaire</i>	5 500	6 000	500
Source: French study of the 1992 cohort of school leavers			

The financial effect on the individual of investing in different levels of qualification becomes clear. The final and perhaps the most rigorous approach to reporting effects on the

individual is to present information on the effect of the reform on a group of individuals compared to a demographically matched group over a period of time. For example:

Table 8: **The measures of success for two active youth labour market programmes**

	Non participants	Programme 1	Programme 2
Earnings 1 year after programme	SEK 37 750	SEK 52 110	SEK 44 120
Earnings 2 years after programme	SEK 89 300	SEK 74 770	SEK 66 700
Employment within 1 year of programme	37 %	29 %	24 %
Employment within 2 year of programme	42 %	41 %	39 %
Study within 1 year of programme	11 %	10 %	5 %
Study within 2 year of programme	12 %	13 %	9 %
Source: Swedish youth labour market studies.			

It is possible to deduce the relative effects of the programmes on individuals who have or have not participated in the alternative programme. A profile of effects on individuals emerges. Thus it is possible to determine a range of ways of measuring effects or outcomes ranging from the purely subjective view of the impact of intervention on an individual to a much more objective view of the impact of intervention on an individual. All of these have their value. In the case of a formative evaluation of a current programme of reform, the immediate views of the participants can help managers to respond positively and increase motivation to learn and to remove barriers to learning ⁽³⁾. Long-term comparative impact data can identify where further investment can be made with optimal value for money for long-term outcomes. Sometimes the proven effectiveness of the benefit of a training programme

carries little weight with potential recruits compared to the 'grass roots', here-and-now opinion (especially if it comes from peers) about content and pedagogy. Therefore, knowledge of the latter is important if recruitment levels are important.

In addition to measuring the effects of VET interventions on the individual, it is also common practice to measure how the individual was motivated to participate and consequently how they translated the intervention into impact. Motivation is a complex area to examine in evaluation studies. The level of motivation does not depend solely on the programme or future action; it depends crucially on context and a multitude of factors important to individuals. Nevertheless, this complexity does not deter evaluators from seeking information about potential motivating features of programmes, as the following examples show.

Table 9: **Why a modern apprenticeship is attractive**

%	Attracted a lot by ...
78	learning real skills in the workplace
76	career prospects
72	good quality training
70	being employed
69	paid to train
63	earning a wage
32	status of modern apprenticeship
32	being able to leave school
Source: UK modern apprenticeship survey	

⁽³⁾ For an example of this see the German PQO study.

The information above shows the effect of characteristics of the programme on individuals.

Table 10: **Motives for abandoning apprenticeship**

(%)

I didn't like study	49.8
I was thrown out	12.9
I would like to study independently	9.1
I was not doing well	5.4
Family reasons	5.0
Financial reasons	3.1
Prefer another qualification	1.0
Did not like schools and work	0.9
I am studying (something else)	6.7
Source: evaluation of Italian Apprenticeship	

The effects on individuals of factors external to the provision can be considered when evaluating effectiveness.

Table 11: **Effect of training on long-term unemployed**

Self reported effect	%
Increased self confidence	55.6
Made new friends	43.3
Sense of accomplishment	43.9
Clarified my objectives	20.0
Better able to occupy time	16.1
Improved level of education	42.2
Obtained qualification	30.0
Qualified for further study	20.0
Qualified for further training	11.7
Motivation for further education	29.4
Motivation for further training	21.7
Helped obtain job	10.0
Source: Irish VTOS evaluation	

To the question how relevant is your employment subject to that of the vocational training programme you attended, almost 3 out of 10 trained employed respondents replied very/quite a lot relevant. Consequently, almost 1 out of 10 trained respondents work at this period of time (almost 10 months after the completion of the vocational training programme) in a job relevant to the subject of the vocational training programme he/she had attended.

Source: Greek evaluation of VET

These last four extracts show information about effects on individuals that reform managers can use to optimise recruitment and retention.

Sometimes the main focus of evaluation is on the principal agents managing the VET infrastructure and the effect on individuals can be lost. For example, the effect on students in training linked to UK work related further education (WRFE) reform was unclear. Colleges based their projected demand for courses on employer data (as the main focus of the reform) and failed to notice that students were becoming more conscious of employment as an outcome of further education (FE) and therefore courses with the highest chances of employment were likely to become more popular. However, most colleges in the reform programme continued to run courses with low employment prospects. It was assumed from the outset that employers were the key subjects of the impact of the reform as they determined the labour market responsiveness of the planning process. The students came lower down the list of priorities for judging effects and some were not well served by the reform.

Determining the layers of impact and the relative priority given to different recipients is important in evaluation design. The objectives of evaluation usually make clear who or what is the main subject of impact. Evidence from the set of selected reforms suggests individuals are often the main recipient and providers and enterprises follow. Larger reforms have goals linked to a better socio-economic situation and, while difficult to measure, effects on these follow on. Returning to the temporal dimension, it is also clear from the reforms studied that enterprises and training providers may be the first to experience the changes associated with the reform but in time these are transformed into effects on individuals. Enterprises and training providers are usually the focus of reform process objectives rather than of objectives linked to outputs or effects. In this study examples exist of where the focus of the evaluation objectives are only on enterprises and training providers as part of the process of reform. Others examples exist where these agencies are omitted from the evaluation because the effect of the reform is only measured as an impact on the individual.

6.4. The organisation

There are three main categories of institutions that are subject to impact studies:

- (a) businesses (as employers of trained people);
- (b) providers of training (including businesses as trainers);
- (c) regional or local government.

There was little information in the selection of reports on the impact of interventions on the latter two categories. This may be because it is assumed that the effect of the VET intervention on the individual is the prime outcome and only process evaluation offers added value for money, in that it helps providers and local officials to understand VET better and to place themselves in a position to modify the process for the best. Stern (2003), sees this development of knowledge as a key purpose of evaluation.

A characteristic of measuring the effect of reforms on enterprises is the difficulty of obtaining feedback on impact. Where information from enterprises is available in reports it usually takes the form of satisfaction data rather than more specific information on the changes to work process and productivity that have arisen as a result of the VET intervention. Questionnaires are usually the preferred means of data gathering, and while this kind of survey technique can reach many enterprises, the more specific type of information about impact is difficult to extract and correlate by means of questionnaire. Some of the reports in this sample show poor response rates to surveys; sometimes the evaluators are concerned about this and signal reservations about the generality of findings (e.g. the UK Union learning fund reform). Response rates are only maintained if the questionnaire is simple and not time consuming to complete, again confounding the collection of impact data. In conclusion, the data available for analysis from enterprises is likely to be shallow in terms of outcomes and effects. An example drawn from a report in this study illustrates this: 'A number of benefits are not quantifiable which makes an assessment of impact of a project more difficult' (Union learning fund – UK).

These findings are confirmed by evidence from research studies that enterprises generally do not evaluate the impact of VET but almost always try to estimate costs (Tessaring, 1999).

Reports indicate that employers were motivated to participate in training activities for reasons such as:

- (a) improving the quality of training;
- (b) staff replacement;
- (c) company growth;

- (d) availability of government funding for training;
- (e) improving the general level of education of the work force;
- (f) developing a more motivated, confident and enthusiastic workforce;
- (g) developing basic literacy and numeracy skills;
- (h) improving the specialist skills base;
- (i) improving communication.

In terms of benefits derived from VET activities, which is taken to mean outcomes of the activities, the following features were also identified by businesses:

- (a) project based work was more effective;
- (b) better industrial relations;
- (c) could lead in the longer term to a workforce that was easier to train.

These headings are largely generalisations of more specific reasons for valuing training ⁽⁴⁾.

As in the section above on the impact on individuals, employers' views are sampled for satisfaction ratings, they are also sampled to identify issues, for example the extent of bureaucratic loading and poor relations with training providers. In many cases they are surveyed simply for information, for example: 'how are trainees supervised?'.

The first stage of extending a typical satisfaction survey is to seek the views of training users about changes. The following example shows how useful this extension can be.

In the German PQO (*Personalqualifizierung in Ostdeutschland*) study enterprises singled out the following outcomes:

Table 12: **Changes in workers as a result of VET**

	Number of enterprises
Have become more qualified	33
Are more motivated	27
Have taken on new tasks	10
Have been promoted	3
Have received better pay	6

The following example, drawn from the UK WRFE reform based on a large sample of enterprises, makes seeking views, rather than hard data on outcomes or effects, a little more useful by providing a comparable analysis of employer views of training courses. Some colleges had participated in a reform to develop training

specifically to match the needs of enterprises in their locality, while other colleges were not involved. The views of enterprises in the regions served by the colleges which were participating in the reform were compared to the views of enterprises in the region where colleges were not involved.

Table 13: **Employers' views on the reform**

Colleges provide training very well/quite well in terms of:	employers' views in regions of the reform	employers' views in other regions
The courses provided	76.1	42.6
Relevance of course content	77.5	39.1
Using appropriate machinery	75.5	34.7
Training for a workplace setting	69.1	34.9
The timing of the courses in the year	80.1	60.3
The mode of attendance	84.9	62.1
Cost effectiveness	84.0	63.7

(%)

⁽⁴⁾ More details on this issue can be found in a synopsis of research into how training impacts on enterprises is given in Tessaring, 1998, pp. 72-77.

This is useful material for those judging the effectiveness of the scheme and could be used for any extension and dissemination of the reform. An extract from the Austrian Fachhochschule reform shows another example of

how satisfaction ratings can be made more scientific. In this case the views of enterprises were sought about the qualities of graduates from the new vocational institutions compared to graduates from traditional universities.

Table 14: *Fachhochschule* graduates compared to graduates from traditional universities

(%)

	Better	Same	Worse
Ability to meet practical work demands	37.3	49.3	13.4
Ability to work in a team	29.9	66.2	3.9
Relevance of qualifications	27.0	59.0	14.0
Introduction phase (length of time spent)	21.3	72.5	6.2
Ability to cope with new tasks (speed)	19.3	66.7	14.0
Creativity in problem-solving	15.6	67.5	16.9
Ability to apply their knowledge	13.0	70.1	16.9
Career chances	7.8	77.9	14.3

Surveys of employers sometimes pinpoint where VET can be improved. Enterprises involved in the German PQO initiative identified three areas:

- consider the demand side of training carefully and customise training accordingly;
- be proactive in marketing their training programme;
- be flexible about when and where to deliver training.

These features were also identified as effects of the UK review of work related learning in colleges. Employers witnessed growth in tailor-made courses and in methods used for identifying demand. They believed colleges had become more flexible.

The German PQO study sought opinion on the characteristics of good training. Good training was rated by employers as better in the following ways:

- breadth of courses;
- exclusion of extraneous skills from courses;
- more appropriate machinery used;
- links with the company's way of working;
- handling of suggestions or complaints.

Only the first two had statistically significant positive outcomes from employers.

In reforms where fundamental changes to the structure of VET are taking place (for example in Spain and Portugal) effects on employers can take on a different focus. The chief area of impact is involvement itself and participation measures are used extensively to judge the effects of reforms. This is as far as assessment of impact on enterprises extends in countries with these large-scale reforms.

In many of the more educationally focused reforms, for example development of qualifications systems, businesses were not considered a direct source of information about impact in the reforms selected. What would be a measure of impact on an enterprise of a school-based reform? It is interesting to think about what useful measures of effects, examples might be: easier recruitment, lower initial training/induction costs, better skills supply or reduction of staff turnover.

In the light of all of the difficulties of gathering employer views of impact, the wish to continue involvement with a VET initiative may be the best indicator of an employers' view that the impact is positive. However, more information would be gained if values were compared for different pathways. For example:

Table 15: **Are you interested in employing more graduates from *Fachhochschule* in the future?**

(%)

	1998	2000
Yes	60.6	65.5
Maybe yes	27.2	21.4
Probably no	12.1	7.1
No	0	6.0
Source: Austrian vocational higher education initiative		

Trade unions have not usually been the subject of impact evaluation, though in the case of the introduction of the Union learning fund in the UK the effect on unions was measured specifically. The reform aimed, amongst other things, to develop the capacity of the union to deliver training and to distribute funding for training. When the unions were asked to define what effects had occurred on building capacity they had diverse views: some stated 'the delivery of learning activities'; others mentioned 'developing credibility' and 'developing sustainable resources such as websites'. Clearly the definition and agreement on likely or desired effects at the stage of programme design would have helped in determining capacity building. The evaluators suggest it is possible to group the diverse kinds of effects for unions under three headings:

- (a) developing resources and expertise;
- (b) securing ownership of the lifelong learning agenda within the union;
- (c) developing union strategy and policy.

6.5. The aggregate level of the economy and society

It is much more difficult to measure the effects of VET interventions on broader society. The links between improved individual performance, improved commercial performance, greater well-being for individuals, their families and their community is hard to establish in quantitative terms and none of the reforms selected have attempted to establish such links. However, some of the reforms, notably the German and Spanish reforms, have had as their central aim the development of the regional economy.

In their report on the impact of human capital on economic growth, Wilson and Briscoe (2003) find that several fields of economic research suggest a link between investment in education and training and general national economic performance. For example they show that micro-economic research suggests '... strong links ... between education productivity and output levels. Although some have questioned the direction of causality and argued that much education simply acts as a screening device to help employers to identify more able individuals, the general consensus seems to be that education does result in higher individual productivity and earnings. On balance, the results suggest a strong and positive causal link between investment in education and training and earnings. This applies both at the level of the individual and when considering the broader social returns to such investments. The evidence seems to suggest substantial social as well as private benefits to this kind of activity. The implication is that what is good for the individual at a micro level is also good for society at large at a macro level.'

They reach a similar conclusion when they consider 'rates of return' literature and the literature on skills and organisational performance.

Green et al. (2003) reach similar conclusions about the immaterial benefits of VET. They also look at the micro level and cautiously affirm that: 'raising educational, skills and training levels is neither a necessary nor sufficient condition of promoting macro-social 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.'

Many of the reforms, notably the Nordic reforms, have had the general well-being of

society at the heart of developments. It will be difficult and will take time to evaluate the effects of IVT on these broad goals and it is likely that other social and economic changes will render such analysis impossible. However, proxy measures for social change are being used and

do provide some impact data. Dropout from upper secondary education is often reported as a proxy measure; the reasoning is that new IVT courses will appeal to young people who might have rejected further non-compulsory general education.

Table 16: **Five years after gaining a qualification how many young people in France are registered as unemployed or inactive?**

(%)

Qualification	Registered unemployed	Inactive
<i>Bac technologique industriel</i>	13	1
<i>Bac professionnel industriel</i>	11	0.5
<i>Bac général</i>	12	6
<i>Bac technologique tertiaire</i>	18	4
<i>Bac professionnel tertiaire</i>	16	3
Source: French study of the 1992 cohort of school leavers.		

Another proxy measure for evaluating the wider impact is the success of individuals

entering work and/or study. This data is often reported as trends, for example:

Table 17: **Destinations on completion of Portuguese IVT programme**

(%)

	1998	1999
Continuing studies	42	29
Employment (active life)	39	41
Other	19	30

In some reforms the aim was to provide a means for the excluded to gain access to education and training so that they might return to work. Clearly there are wider benefits for society (as well as for the individual and their family) for reducing the number of those dependent on benefit payments. These proxy measures of wider impact are also

reported. For example, the Irish report on VTOS attempts to show the effect of education and training on unemployment. The effect on unemployment of VTOS participation is not reported against that of non-participation but the evaluators focus instead on the reduced probability of unemployment with higher levels of qualification.

Table 18: **Labour-market situation immediately following VTOS by certification achievements**

(%)

Qualification	unemployed	in education	in job
5 or more subject leaving certification	80	20	0
4 or less subject leaving certification	70.6	17.6	8.8
Portfolio of achievements	76.9	23.1	0
Vocational qualification	68.8	3.1	28.1
No certification	64.6	6.2	16.9

The ease of entry into the labour market after training is clearly a key output measure. In some reforms, for example the Portuguese IVT development and the Irish VTOS reform, there is a requirement that training organisations consult with local employment agencies and employers before determining the training provision they provide. Unfortunately there is no comparative data available for drawing conclusions about the impact of this requirement.

Finally, the impact on broader society of reforms which have a decentralisation of control at their centre (for example the Italian, Spanish and Swedish reforms) is important. No evaluation of this kind of impact is reported nor is there evidence of the intention to attempt to measure this type of impact. However, there are likely to be significant impacts arising from the effects of

local decision-making and greater opportunity to focus on issues more important locally than they are nationally. The driving forces for these reforms are often located in the notion of maximising the effect of VET and minimising bureaucracy.

Labour-market dynamics, training infrastructure, qualifications systems, funding schemes, societal pressures, cultural values are all deeply inter-related and to look for effects in one dimension may not be the best way of informing improvements for individuals and businesses.

Having studied and analysed the set of reports in depth, there are some general points which emerge which could help with improving evaluations in the sense of making them more impact oriented. These points are discussed in the next chapter.

7. Towards better evaluation of reforms

In general terms impact should be measured with three parameters in mind:

- (a) qualitative measures: measures that locate the intervention in a wider context of the issue being addressed and what theory we can draw on to illuminate the changes in the position of stakeholders;
- (b) quantitative measures: the depth and breadth of the impact of the intervention on potential stakeholders in relation to those of comparable position that are not participating;
- (c) temporal measures: indication of impact over a longer timescale to see effects that may result from the initial intervention.

The analysis of the evaluations of VET reforms has suggested that there are some steps that can be taken towards securing a better evidence base for their impact. These steps do not represent fresh thinking about impact analysis; they simply emerge from the reports in response to the question ‘how could we discover more about the outcomes of reforms when we evaluate them?’.

In setting up an impact evaluation the following considerations might improve the measurement of the impact of a reform:

- (a) identify what is driving the reform; why is the reform happening? This examination can clarify the need for impact analysis; will the data on impact affect the forces for change? Will these forces continue regardless of feedback?
- (b) clarify the purpose of the evaluation programme. Is the identification of outcomes on individuals, enterprises and wider communities adequately represented in the purpose of the evaluation?

- (c) ensure the objectives of the evaluation programme cover the desired outcomes of the goals of the reform. Check that the data which is likely to be generated by the evaluation can be analysed in a way that fully meets the objectives of the evaluation programme and covers the desired outcomes of the reform;
- (d) consider methodology and use opportunities to make the process as rigorous as possible. This does not necessarily involve developing quantitative approaches but it does mean considering baseline measurements and timelines carefully;
- (e) if possible, analyse the effects of variables that cannot be controlled, for example the changes in the vigour of the labour market or changes in attitude towards certain working methods;
- (f) where resources allow, use survey methods that are sensitive to context. These usually involve using semi-structured schedules in face-to-face interviews, telephone surveys, focus groups or on-line adaptive survey techniques;
- (g) consider how the results will be presented to sponsoring organisations and highlight the need for an extended timescale for true impact statements to emerge and to take account of the tentative nature of interim impact-related findings.

The following example illustrates how the process of breaking down the objectives of a reform and transforming them into a plan for evaluation can raise the level of impact measurement.

From aims to enquiries

It may be helpful if the following process is used to identify areas where impact measures may be located and profitably measured. It involves resolving the aims and objectives of reform into lower level questions for evaluation. In reporting impact the answers to these lower level questions can be reported and aggregated to determine a response to the main questions.

Main reform objective: to introduce employer-led apprenticeship

One of the objectives for evaluation: to determine the response from employers

A qualitative response and a quantitative response are sought – taking the latter first:

How many employers were involved?

How many were eligible?
 Of those involved what are the prominent sectors, what size of companies?
 Of those not involved what are the prominent sectors, what size of companies?
 Regional differences?
 How many are planning to be involved in the future?
 Etc.

For the qualitative response:

Of those involved – satisfaction ratings, including longer-term plans. Changes to non-apprenticeship training plans, changes to recruitment practice and policy.

Of those not involved – reasons why not, plans for the future, potential inducements required.

The impact of a reform occurs at several levels; evaluation programmes might consider the primary impact on individuals and institutions, secondary effects on user organisations and communities, and tertiary effects on society and the economy.

A key step in enhancing impact measures is to consider the timing of data gathering from stakeholders. This usually takes place in stages: during the reform, immediately afterwards, one year afterwards and three to five years afterwards.

8. Summary and conclusions

The 19 reports in this meta-study provide an overview of the kinds of national reforms to VET that have taken place over the last 10 years. The selection of reports (the main focus being impact study) may have distorted the review but it seems clear that VET is now less centrally determined than it once was and that its educational role is now wider in terms of participation, content and assessment.

Key dimensions of impact evaluations have been highlighted: the ways in which data is collected can facilitate more scientific analyses; in-depth study of the aims of reforms and appropriate objectives for evaluations; careful consideration of levels of impact; and the timing of measurements is important.

Common evaluation criteria have been developed which have the potential to make studies of this kind more systematic. The criteria have proven to be useful in highlighting aspects of evaluation to be considered in setting up a programme. As a tool for analysing existing reports they functioned well. However, they were limited in their capacity to create comparative analyses by the varied nature of the reports in this study. The criteria used to evaluate reports are worthy of wider dissemination and are now ready for application in other contexts.

This research has shown that evaluation practice in the context of VET across EU countries is variable, ranging from a wish to integrate evaluation into the infrastructure of reform to practices which minimise investment in evaluation. Evaluation of impact is also generally restricted to satisfaction surveys of current experience of a reform programme and they rarely investigate the long-term effects of programmes. There are strong indications that there are differences in the way countries value evaluation. Further work is necessary to elaborate on this point but experts have reported a dearth of evaluation, for example, in Germany and in Greece. There seem to be thorough evaluative approaches linked to reforms in Portugal and Spain. In Finland and in Sweden we see evaluation tightly entwined with the reform process.

It is possible that this variability may also be related in some countries to the availability of people who are confident and experienced in both VET policy planning and delivery and in evaluation methodologies. Cedefop has a role in developing expertise in the impact evaluation of VET interventions. Dissemination of good practice is the key and the publication of the third research report will be an important step.

There may be circumstances where cross-national studies of impact evaluation are sought. The experience of this study indicates that a panel of researchers would probably be the best way of carrying out such studies. This would have the advantage of sharing expertise as well as ensuring sensitivity to any national context by using experts from different countries on the panel.

The examples in the previous chapter show that it is relatively easy and inexpensive to improve basic satisfaction surveys to provide more information on effects. More work can be done to construct guidance for those embarking on evaluation design applied to VET. This guidance must be sensitive to the political nature of evaluation, to the socioeconomic context in which it is conducted and to the range of stakeholders with an interest in VET reform.

Such an improvement in the volume and quality of impact analyses will possibly influence those who commission evaluations to be more confident that impact analysis is reliable, valid and useful. Until impact analysis can be relied on for what it can offer it will not draw the investment it deserves and will not be used as much as it should. The strength of impact analysis is also its weakness in that it is a long-term evaluation of a reform and reports on whether it has worked or has not worked. Commissioners of process evaluation are confident about it because it is short term and yields results that give a sense of good housekeeping and uses open, democratic processes.

The reports studied contained references to the driving forces that create VET reform. However, there is little detail provided about how

these forces operate and how feedback from evaluation programmes might intersect with them. One of the main reasons for conducting impact evaluations is that this intersection is facilitated. Research might usefully be conducted on the dynamics of this process and could begin with a review of the literature describing the driving forces behind the reform of VET.

The findings of this study are soundly based in the evidence from the sources used. However, it is clear that the research conducted could be extended. The current study is limited by such factors as the availability of reports and it is possible that some useful reports of reforms have not been included in this study. The study would also benefit from a more extended understanding

of national contexts, particularly about the key developments in countries undertaking reform of VET. It may prove useful to study the evolution of a VET system and trace the drivers of change. In so doing the informal process of evaluating provision and making incremental changes as a result should become clear.

The meta-study methodology used in this study is limited by what is written and published about a reform and its evaluation strategy. It is likely that much more evaluative evidence exists in the minds of reform managers and evaluators. Access to this information demands careful strategic planning, face-to-face interviews and assurances as to the purposes of the wider meta-study process.

List of abbreviations

CVT	Continuing vocational training
IVT	Initial vocational training
PQO	Personalqualifizierung in Ostdeutschland
QCA	Qualifications and Curriculum Authority
QVE	Qualified vocational education
VTOS	Vocational training opportunities scheme
WRFE	Work related further education (programme)

Annex: common evaluation criteria

Clarification of purpose of these criteria

We are looking for criteria we can apply as we read evaluations of programmes that come under our remit. These will be fairly large-scale evaluations completed by reputable individuals and organisations.

While we are concerned about 'impact' evaluation, we will take account of 'process' evaluation if we are to make useful conclusions about evaluating impact and the design of reform programmes. Hence the breadth of territory covered by the criteria below.

The criteria are grouped by chronology of project implementation. They are further grouped by the following characteristics of evaluation programmes:

- (a) context: what external influences are likely to influence?
- (b) goals: how are the aims, objectives, milestones described?
- (c) planning: what is discussed in anticipation of what might happen?
- (d) management: what is done to make the programme function effectively?
- (e) drivers: who is shaping the evaluation activity?
- (f) impact: what is the effect of the evaluation?

Before the implementation of the programme

Context

Reform setting

Was the background (political, social, economic, cultural) for the reform made clear? How much attention was given to setting when designing the evaluation?

Policy relevance

Was the nature of any explicit link between national economic goals and the reform (e.g.

skills needs, workforce development and partnership) made clear? Was consistency of reforms with other policies evaluated?

Policy description

Was a description of national policy used to underpin the reform and its evaluation? Was it intended to measure impact? If there was no policy description what was the purpose of the evaluation?

Goals

Reform aims

Was the aim of the reform included? What was to change (e.g. learning, institutions, frameworks, and certification)? What was to be maintained? Was this the principal focus of the evaluation programme? Were there clear targets for the reform?

Purposes of evaluation

Was the purpose(s) of evaluation made clear? For example to inform decision-making (government, employers, providers, and trainees), improving reform management, improving resource allocation, improving the quality of programmes, a mechanism for stimulating debates, a process of bringing reform designers and managers to account.

Theoretical basis

Did the reform have an explicit link between hoped-for outcomes and a theoretical basis for expecting the interventions to lead to these outcomes?

Planning

Baseline for analysis

Was the evaluation set up in such a way that a baseline for measuring the effects of interventions was established? Was this baseline established and agreed amongst stakeholders before the start of the programme?

Management

Management structure

Was the management infrastructure of the reform discussed? Who was consulted? Did the evaluation look at this?

Donor commitment

Is there a declaration of intent and continued commitment to the reform from any donor authority? Did the evaluation investigate it?

Risk management

Were the risks associated with the reform made explicit? What was done to reduce risk? How was risk management organised? Did the evaluation include programme risks?

Legal embedding

Is there evidence in the evaluation that legal arrangements were considered as a preparatory step towards implementing the reform?

Programme handbook

Was there a publication describing the reform? Who was it aimed at? How did any donor institution contribute to this publication? Was the manual evaluated for content and usefulness to users?

Preparatory studies

Were feasibility studies carried out or ex ante evaluations planned? Did the evaluation look at the quality of these studies?

Drivers

Reform actors

Was there evidence of pressure for change? Who was opposed to it? How did they act? Was there evidence of involvement from policy level actors? Similarly was there evidence of bottom-up demand for reform? Was a stakeholder analysis part of the evaluation?

Impact

User impact

Was it planned to make interventions that could be measured by changes in opportunity and practice amongst end users (learners, communities, organisations, and countries)? How was the evaluation organised to measure impact?

During implementation

Context

Familiarisation

Did the proposed reform become widely known? Who discussed it? Did evaluation consider dissemination throughout the duration of programme? If so what issues were discussed and why?

Goals

Stability

Did the goals of the reform change during the implementation? If so did the evaluation identify the influences that were driving the change?

Planning

Formative influence

Is there evidence of change in intervention programmes as a result of information arising from the evaluation process or intermediate outcomes of the reform? Was evaluation ever present during discussions of the reform?

Complementary activities

Is there evidence of organisations at national, regional and local levels working jointly towards common targets in a complementary top-down and bottom-up approach, and were they involved in the evaluation?

Management

Transitional arrangements

Were there transitional arrangements in place between existing VET systems and those introduced as a result of reform during the period of reform? How were these introduced, managed and dismantled? Did evaluation consider this?

Piloting arrangements

Is there evidence that any pilot phase was useful because it linked strongly to systemic change, and was effectively disseminated? Were issues of mainstreaming and sustainability addressed during the expansion phase, and did evaluation consider these?

Drivers

Stakeholder motivation

Were people and organisations incentivised to participate? Did it work? How was this incentivisation differentiated and assessed?

Stakeholder groups

Were key groups participating in VET involved? For example, those in pre employment activities, those upgrading their skills used in their job, those retraining in another area of work, those requiring VET to improve employability.

Stakeholder activities

Was the involvement of individuals and key groups monitored? How were they used as indicators of success or failure in programme and the evaluation?

Impact

Timely information

Was there a means of summarising evaluation information and regularly feeding in to the reform management activity? How is this done?

After the programme had reached maturity, been completed or was terminated

Context

Programme context

Was the political, social, economic and cultural influence considered important for the success of the reform, and how did evaluation take context into account?

EU priorities

Were the EU-wide common themes for VET (such as improving access to VET and increasing participation) addressed in the programme, and the evaluation?

Goals

Duration of outcomes

Was there a plan to look at long-term effects of

interventions? How sustainable is the reform? Was this evaluated?

Planning

Dissemination phasing

Is there evidence of the dissemination process culminating at this point at the outset of the activity?

Management

Reform organisation

Did evaluation consider the appropriateness of the basic structure of the reform and its interventions (e.g. legal, structural, management, funding)?

Programme timing

Was a timescale specified and did the reform process meet its time-related goals?

Programme budget

Were cost considerations (of reforms) included in the evaluation specification? What were the intended costs?

Monitoring tools

Were monitoring and evaluation tools in place to identify, take account of and measure the impact of policy changes during the course of the reform and did evaluation analyse them?

Programme flexibility

Was the programme adjusted to take account of changing financial, social and political contexts? Did the evaluation programme seek evidence for this?

Drivers

Social partners

Was there a plan for full engagement of social partners? Were they involved in the evaluation?

Programme spillover

Is there evidence that the programme and evaluation practices (introduced from the EU) resulted in activity in a Member State and a programme of national action?

Impact

Personal impact

Is there evidence that the intervention has had an impact on individuals for job change, learning progression, and reward?

Institutional impact

Is there evidence that the intervention has had an impact on institutions in terms of behavioural transfer of learned skills into results?

Societal impact

Was the impact of interventions on communities measured? For example, families and social organisations.

Systemic impact

Is there evidence of a systemic link between the outcomes of the interventions included in the reform and policy development, and how was it measured?

The evaluation process: some overarching criteria

Development Process

Was there evidence that the evaluation was understood as a development process, which aims at capacity and institution building, and creates self evaluation skills amongst beneficiaries? Was the evaluation integrated into the reform process?

Culture

Were the perceptions of major stakeholders (EU funders, national government, major actors, and social partners, evaluators themselves) made explicit?

Autonomy of evaluators

Was there evidence that the evaluators had some choice in the style and direction of the evaluation and in the choice of tools? Is the evaluating institution independent of the reform?

Coordination of evaluation projects

Was the evaluation programme made of separately commissioned elements? If so how well coordinated was the whole programme?

Method

Were practical constraints on the evaluation process discussed? Was there evidence of methodological weaknesses in the evaluation programme – e.g. sampling, instrument design, time and constraints? Did the impact analysis include a control group? If so was the process experimental or quasi-experimental? Was there a combination of qualitative and quantitative assessment of impact?

Completeness

Is there evidence of the whole process of reform being fully evaluated? Were there shortfalls in evaluation effort in specific phases?

Indicators

Is there evidence of use of quantitative and qualitative indicators being developed for the evaluation process? If there were no indicators (developed by, for example, benchmarking and target setting) what effect has this had?

Evaluation skills

Was the experience of the evaluation team discussed? Were the methods and mechanisms understood by stakeholder organisations?

Evaluation censorship

Was there any evidence of evaluation findings and reports being shelved or not disseminated?

Evaluation budget

Was the size of the evaluation budget appropriate in relation to the activity as a whole (as a rule, 5 % of budget for small projects, and 3 % for large programmes should be made available)?

Reporting formats

Is there evidence that the dissemination materials are being targeted to different audiences as opposed to a one-size-fits-all policy?

Evaluation dissemination

Are evaluation results being made available to the project partners and to all other member countries?

References

- Angeli, F. *Il nuovo apprendistato. Rapporto 1999*. Rome: ISFOL, 1999.
- Antill, M. et al. *An evaluation of the Union learning fund in year 3*. London: DfES, 2001 (York Consulting Report).
- Baker, J. *Evaluating the impact of development projects on poverty: a handbook for practitioners*. New York: World Bank, 2000.
- Baker, S.; Taylor, J. Effects of career progression interventions: a meta analysis. *Career Development Quarterly*, 1998, No 46, p. 376-385.
- Barrett, A. Methodological introduction. In: Elson-Rogers, S. (ed.) *Approaches and obstacles to the evaluation of investment in continuing vocational training: discussion and case studies from six Member States of the European Union*. Luxembourg: Office for Official Publications of the European Communities, 1998, Part I, Chapter 2 (Cedefop Panorama series, 78).
- Borghans, L.; Heijke, H. *Does the Dutch adult and vocational education act promote the macroefficiency of vocational education?* 2002 (unpublished).
- Coleman, N.; Williams, J. *Evaluation of modern apprenticeships: 1998 survey of young people*. Sheffield: Department for Employment, 1998 (Research Report, No 93).
- Cutter, J. et al. *A second evaluation of the Union learning fund*. London: DfEE, 2000 (York Consulting Report).
- De Bruijn, E.; Doets, C.; van Esch, W. Flexibility in secondary vocational education. *Pedagogische Studien*, 2001, Vol. 78, No 6.
- Department of Employment. *Technical and vocational initiative report*. Sheffield: Department of Employment, 1994.
- Descy, P.; Tessaring, M. (eds) *Training in Europe: second report on vocational training research in Europe 2000: background report*. Luxembourg: Office for Official Publications of the European Communities, 2001 (Cedefop Reference series, 3008).
- Economic Research Services Ltd. *Evaluation of modern apprenticeships: survey of employers*. Sheffield: Department for Employment, 1998 (Research Report, No 94).
- Ekstrom, E. *The value of a third year in upper secondary education: evidence from a piloting scheme*. Uppsala: IFUA – Institute for labour market policy evaluation, 2002 (Working paper 2001: 23).
- European Commission, Monitoring and Evaluation Unit, Employment and Social Affairs. *Guidelines for systems of monitoring and evaluation of ESF assistance in the period 2000-2006*, Brussels: European Commission, 1999.
- Felstead, A.; Unwin, L. *Funding systems and their impact on skills*. London: Department for education and Employment, 1999 (Skills task force research paper, No 11).
- Finnish Ministry of Education. *Valtioneuvoston selonteko eduskunalle uuisen koululakien vaikutuksista ja laessa asetettujen tavoitteiden toteutumisesta*. Helsinki: Ministry of Education, 2000.
- Forcem, Fundacion para la Formacion Continua. *Memoria balance – I acuerdo nacional de formacion continua 1993-1996*. Madrid: Forcem, 1999.
- Freundlinger, A.; Grillitsch, S.; Schmid, K. *Evaluation neuer Lehrberufe*. Vienna: Institut fuer Bildungsforschung der Wirtschaft – Schriftenreihe/Forschungsbericht, 2000.
- Green, A.; Preston, J.; Malmberg, L. *Immaterial benefits of education, training and skills at a macro level*. In: Descy, P.; Tessaring, M. (eds) *Third research report: background report*. Luxembourg: Office for Official Publications of the European Communities, 2003 (Cedefop Reference series).
- Griffith, J. An approach to evaluating school-to-work initiatives: post secondary activities of high school graduates of work-based learning. *Journal of vocational education and training*, 2001, Vol. 53, No 1, p. 37-60.
- Grubb, N.; Ryan, P. *The roles of evaluation for VET: plain talk on the field of dreams*. Geneva: International Labour Organisation, 1999.
- Irish Ministry of Education. *Report of the national evaluation of the leaving certificate applied*. Dublin: Department of Education and Science, 2002.

- ISFOL. *Il nuovo apprendistato e la formazione – il monitoraggio dei progetti sperimentali*. Roma: ISFOL – Istituto per lo sviluppo della formazione professionale dei Lavoratori, 2000 (Pragma sr-1).
- Kaasbøl, E.; Pedersen, F.; Størner, T. *Eleverne og eud-reformen: oplevelser af forsøg med grundforløb og hovedforløb i 2000*. Copenhagen: Undervisningsministeriets forlag, 2001 (Uddannelsesstyrelsens temahafteserie, No 9).
- Kath, F. The funding of vocational education and training: an international comparison of objectives and impact. *European Journal of Vocational Training*, 1999, Vol. 17, p. 38-44.
- Kirkpatrick, D. *Evaluating training programmes, the four levels*. San Francisco: Berrett-Koehler, 1998.
- Klein, T. *Les bénéficiaires du contrat de qualification entrés à l'automne 1994: les enseignements de 4 vagues de panel*. Paris: DARES, Ministère de l'Emploi et de la Solidarité, 2002.
- Larsson, L. *Evaluation of Swedish youth labour-market programmes*. Uppsala: Department of Economics, 2000 (Working paper series, 2000-6).
- Lechner, M. An evaluation of public-sector-sponsored continuous vocational training programs in East Germany. *Journal of human resources*, 2000, Vol. 35, No 2, p. 347-375.
- Martin, S. A critical analysis of the evaluation of academic performance in Spanish educational reforms. *European Journal of Education*, 2001, Vol. 36, No 1, p. 77-90.
- Metcalfe, H.; Dawson, C. *Evaluation of the work-related further education programme*. Sheffield: Employment Department, 1995. (Research Series, No 55).
- Ministério da Educação e para a Qualificação e o Emprego. Despacho conjunto No 123/97 de 16/6/1997. In *Diário da República II série do 7/7/1997*. Lisbon. The 1998 version is also used in this study. The official Ministries summaries from 1997 and 1998.
- Montavavao e Silva, J.; Santos Silva, A; Protes da Fonseca, J. M. *Avaliação do sistema das escolas profissionais*. Lisbon: Ministry of Education, 1996.
- National Agency for Education. *Reformeringen av gymnasieskolan – en sammanfattande analys*. Stockholm: National Agency for Education, 2000 (Report No 187).
- Neubert, R.; Steinborn, H-C. *Personalqualifizierung in den neuen Bundesländern, Prozessanalyse und Evaluierung (PQO)*. Bonn: BIBB, 1999.
- Nielsen, S. P. *Vocational education and training in Denmark*. Introductory note prepared at the request of the European Centre for the Development of Vocational Training (Cedefop), 1997. Available from Internet: http://www2.training-village.gr/etv/vetsystems/report/dk_en.asp [cited 4/9/2003].
- Nieuwenhuis, L.; Shapiro, H. Evaluating systems' change policies in VET: learning from Danish and Dutch cases. In: Descy, P.; Tessaring, M. (eds) *Third research report: background report*. Luxembourg: Office for Official Publications of the European Communities, 2003 (Cedefop Reference series).
- Nijhof, W. J. (ed.) *Evaluation of the Dutch VET system*. Den Bosch: CINOP (unpublished collection of articles)
- Plewis, I. *Evaluating the benefits of lifelong learning*. London: University of Education, Institute of Education, 2001.
- Polder, K. *Evaluation of an act: a case study of the Netherlands*. Amsterdam: University of Amsterdam, 2001.
- Richini, P. *Lifelong learning in Italy: the extent to which vocational education and training policy is nurturing lifelong learning in Italy*. Luxembourg: Office for Official Publications of the European Communities, 2002. (Cedefop Panorama, 43).
- Shapiro, H. et al. *Erferinger fra forsøg med eud refoemen – grundforløb og hovedforløb*. Copenhagen: Danish Ministry of Education, 2001.
- Stern, E. Philosophies and types of evaluation research. In: Descy, P.; Tessaring, M. (eds) *Third research report: background report*. Luxembourg: Office for Official Publications of the European Communities, 2003 (Cedefop Reference series).
- Summa, H. The role of evaluation in European Politics. *Internationales Jahrbuch der Erwachsenenbildung*, 2000, p. 49.
- Tessaring, M. *Training for a changing society: a report on current vocational education and training research in Europe*. Second edition. Luxembourg: Office for Official Publications of the European Communities, 1999 (Reference Document).

- Thireau, V. *Ouverture de l'accès au baccalauréat et insertion des jeunes: enseignements et pistes ouvertes par Génération 92*. Marseille: Céreq, 2001.
- VFA – Valter Fissamer and Associates Ltd (for Greek Ministry of Labour and Social Affairs). *Evaluation of the labour-market policies and assessment of the influence of European Employment Strategy in Greece during the period 1997-2001 – Final report*. Athens: VFA, 2002.
- Wilson, R. A.; Briscoe, G. The impact of human capital on economic growth: a review. In: Descy, P.; Tessaring, M. (eds) *Third research report: background report*. Luxembourg: Office for Official Publications of the European Communities, 2003 (Cedefop Reference series).
- Wimmer, P. *Die ersten Fachhochschul-Absolventinnen am Arbeitsmarkt*. Vienna: Institut fuer Bildungsforschung der Wirtschaft-Schriftenreihe/Forschungsbericht, 2000.
- World Bank. *Active labor market programs: impact evaluation techniques*, 2002. Available from Internet: <http://wbIn0018.worldbank.org/HDNet/hddocs.nsf/0/02BF4C7B59DE32C985256BF30067A80F?OpenDocument> [cited 3/9/2003].
- WRC Social and economic consultants Ltd. *An evaluation of the vocational training opportunities scheme*. Dublin: Department of Education, 1994.



Evaluating systems reform in vocational education and training: learning from Danish and Dutch cases

Loek Nieuwenhuis, Hanne Shapiro

Abstract

This paper refers to the use of evaluation in the policy of VET system change. Recent reforms of the VET systems in Denmark and the Netherlands are used as case studies to underpin the analysis.

Changing VET systems is a complex issue. Social systems such as VET are multi-layered and to be effective, system change should be consistent at all levels. Furthermore system change is not a one off event, but a continuous process of overlapping developments. Within such complex change policies, evaluation can be used in different ways, from more positivistic approaches to participatory approaches. Positivistic approaches are more compatible with controlling policies whereas participatory approaches fit learning policies better.

Two reforms are described extensively, particularly concerning the way evaluation and research has been used by the reform policy. Reform 2000, the Danish reform of the VET system since 1999, is a good example of the participatory approach. This reform tries to combine national debates and policies with local and regional experiments with new pedagogical tools and steering mechanisms; evaluation and research were used in a learning methodology. The introduction of WEB, the 1996 Dutch Act on VET, is evaluated in an ex post way in 2001, checking whether the WEB objectives have been reached after five years of implementation. The use of evaluation in the WEB case has to be assessed as rather positivistic. The use of evaluation as an instrument for policy learning has been successful in Denmark, but failed around WEB.

The emergent knowledge based economy demands VET policies built on a contingent development strategy in which evaluation and research serve as continuous feedback mechanisms. Adaptive policies imply a learning approach, allowing concurrent adaptation, instead of a prescriptive approach, in which the solution is presented as a blue-print. The development of a culture of learning VET policy is still in its infancy.



Table of contents

1. Governance of change in vocational education and training	57
2. Relationships between research, policy and practice	60
3. Educational policy research: setting the stage	63
4. An analytical framework for case studies	65
5. The Danish case: VET Reform 2000	66
5.1. Policy design in a climate of change	66
5.2. Preparatory phase	67
5.3. Creation of an environment of change	67
5.4. Main new elements in the reform	68
5.5. The evaluation of Reform 2000 in Denmark	69
5.5.1. First evaluation: November 1999 to February 2000	69
5.5.1.1. Approach taken	70
5.5.1.2. First findings	70
5.5.1.3. The didactical framework	71
5.5.1.4. The new student role and the contact teacher	71
5.5.1.5. Dissemination of findings	71
5.5.2. Second evaluation: October 2000 to February 2001	71
5.5.2.1. School organisational initiatives	72
5.5.2.2. Variations in strategies with different student groups	72
5.5.2.3. Project pedagogy and assessment	73
5.5.2.4. Teachers' involvement in the change processes	73
5.5.2.5. Instruments for pedagogical innovation	73
5.5.2.6. Roles and governance structures	74
5.5.2.7. The pedagogical space	74
5.5.2.8. Roles and instruments	74
5.5.2.9. The contact teacher	74
5.5.2.10. The logbook and personal educational plan	75
5.5.2.11. Apprentice companies' knowledge of and experience with the VET reform	75
5.6. Final comments on the use of evaluation in Reform 2000	75
6. The Dutch case: implementation of WEB-1996	77
6.1. A short history of changes in Dutch VET	77
6.2. Features of WEB-1996	78
6.3. Developing evaluation criteria for VET-systems change	79
6.4. Results and use of the evaluation of WEB-1996	80

7. Comparisons and conclusions	83
7.1. Two approaches to systems change	83
7.2. Implications for pedagogy and didactics in VET	84
7.3. Towards general European conclusions	86
List of abbreviations	88
References	89

List of tables and figures

Tables

Table 1: Basic beliefs of inquiry paradigms	61
Table 2: Setting a new context	66
Table 3: Evaluation issues, teams, reports and conclusions	81

Figures

Figure 1: A framework for analysing educational change in skilling systems	57
Figure 2: The legislative framework	69
Figure 3: Policy instruments	84
Figure 4: Linear and interactive innovation policies	85

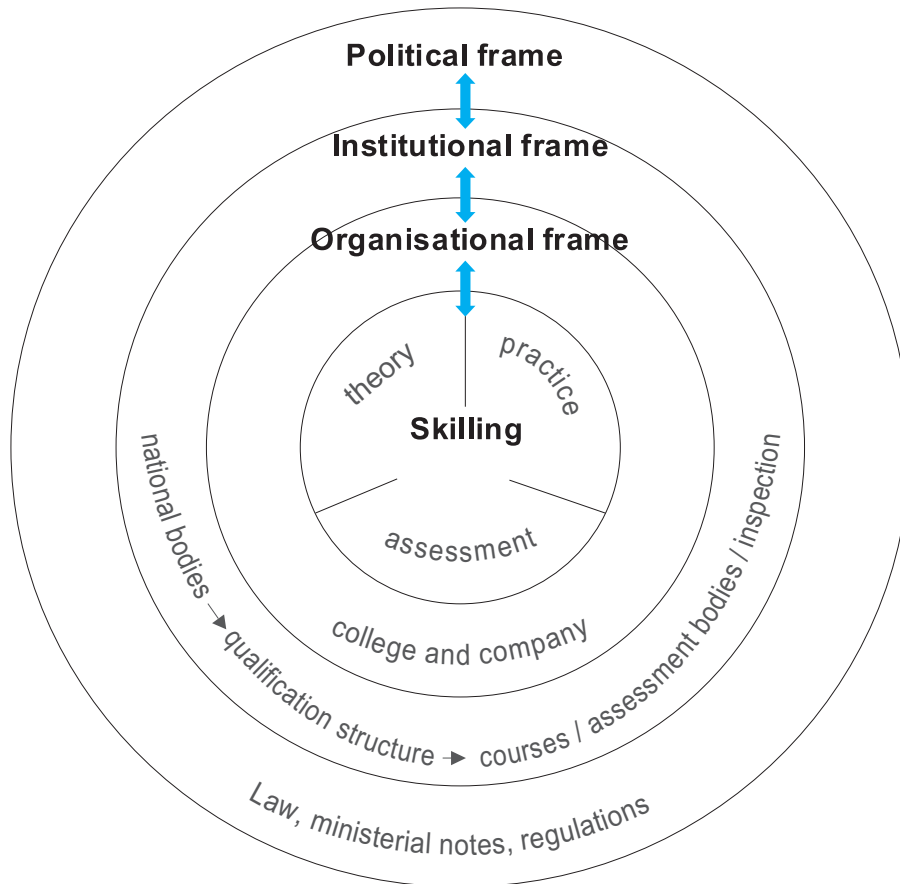
1. Governance of change in vocational education and training

Vocational education and training (VET) is a complex social system because it is located on the edge of two basic human activities: learning and working (Ellström, 1999). For both activities, systems have to be built at local, national and European levels. Within VET, these systems are intertwined in many ways. Changing VET is, therefore, a long term enterprise, requiring interaction and debate between the different interest groups in the system.

According to Edquist and Johnson (1997), social systems are specific creations of institutions and organisations. Social systems are designed incrementally to reach societal goals:

- (a) institutions can be defined as sets of common habits, routines, rules or laws that regulate the relations and interactions between individuals and groups. Functions of institutions are, for example, the provision of information and the reduction of uncertainty; managing conflicts and cooperation; the provision of incentives; the channelling of resources;
- (b) organisations are formal structures with an explicit purpose; they are consciously created and they are players and actors in the system.

Figure 1: A framework for analysing educational change in skilling systems



VET is a social system for which, in each country and/or economic sector, a specific set of institutions and organisations has been developed over time. Governments, business and industry, unions and educational organisations have built an institutional system for VET, which is rooted in social, cultural, and economic patterns. Because of these roots, VET institutions are difficult to change and sometimes are even obstacles to innovation. Changing VET is not only about the educational system, but also of the socio-economic system and of cultural traditions. Examples of VET institutions are: laws on education and labour; public-private arrangements; training funds; collective labour agreements; pathways to becoming skilled; qualifications and wages; occupational identity; training traditions.

In theory, organisations are easier to change than institutions, but organisations depend on the institutional system. Thus, technical-rational arguments (e.g. instructional science for VET) for a systems change are not convincing if they are not compatible with the institutional arrangement. This is an important observation for understanding the design and restructuring of VET systems. Policies for change cannot be organised top-down (Bashkar in Sanderson, 2000). Changing VET and its evaluation has to be directed simultaneously at all the levels involved. Interactivity and consistency between the different layers are the main requirements for effective change. The government is one of the players in a complex policy system. Policy, intermediary structures, colleges, companies and teachers should interact in the change process; managing this process is like directing a large orchestra: if one party is out of tune, the whole performance is endangered.

Systems change is not a one off event. Most social systems change is a continuous process of incremental development, sometimes combined with 'earthquakes' (dissipative systems: absorbing a lot of change-impulses without any change; then disrupting in a large change; Sanderson, 2000); changes are often irreversible: systems change has evolutionary characteristics. Social systems are embedded: VET sits between economy, education and social affairs (inclusion policies). Changing systems takes time, especially because of institutional embeddedness. Reasons for change can be endogenous and

exogenous. Systems change is therefore complex and chaotic because of its multi-layer, multi-actor and multi-purpose character.

In Europe, all states are seeking to make their VET systems more flexible. Old systems were built on stable economies and labour-market institutions (e.g. Germany and the Netherlands) or were built on low-skill equilibrium (e.g. UK; Finegold, 1991). In both cases, powerful economic forces urge that VET systems be upgraded and made more flexible; globalisation of markets, high-speed technological innovation and information and communication technology (ICT) development are examples of these forces. The permanent development of and change in the National vocational qualification (NVQ) system in England; the reform of the dual system in Germany; the shaping and reshaping of VET in Eastern European countries; new legislation in the Netherlands; changes in the Italian VET system: there is no VET system that is not under revision. This revision is not merely a technical debate; societal and political issues are at stake (Coffield, 2001; Nijhof, 2001). Everywhere the search is on for a new balance:

- (a) between initial VET and lifelong learning: the old infrastructure is built on the delivery of initial VET; the new one should be able to deliver support for lifelong learning;
- (b) between traditional occupations and flexible qualifications: old-fashioned occupations are slowly disappearing but the institutions are still built on the traditional occupational structure and there is no room in the arena for new occupations;
- (c) between school-based learning and qualification through work experience: learning steered by attainment targets is to be replaced by learning within changing communities of practice while instruction and training is a proven powerful educational tool;
- (d) between social demands and economic markets: should the student be adapted to labour-market demands or can the worker shape his own economical context?
- (e) between employment and entrepreneurship: is the 'new economy' for employed workers or for entrepreneurial workers?

Policy-makers and other stakeholders try to develop these areas, but in each case the institutional arrangement of VET is at stake. The old

regulations and appointments and the social-economic meaning of VET are changing. The design of new VET means the reinvention of VET-institutions but because of the sociocultural roots of institutions this will take time and be a struggle. Stakeholders (like community colleges, employers, trade unions and the government) should reposition themselves towards the newly developing VET systems.

Crouch et al. (1999) offer a sceptical view of the feasibility of a high-skills strategy in western economies. Analysing skill creation systems in seven OECD countries, they conclude that both state-based and market-oriented strategies are doomed to failure. State-based VET will suffer from low responsiveness and innovation, whereas market-led VET will not be able to up-skill the majority of the workforce. Thus each country should consider this view and rearrange their VET system. Crouch et al. (1999) 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 system.

In most European countries VET systems are under reconstruction. Researching these system changes can be done best from an evaluating perspective. Do the efforts for improvement have the targeted impact on the supply of socially and economically requested skills? In the discussions, two evaluative perspectives emerged, a system perspective (how do VET systems meet performance criteria?) and a constructivist perspective where the main emphasis is on generating learning among the actors that can sustain implementation of change. Changing 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 targets of systems change should be defined. There is no one right way for VET, although a European approach suggests a convergence of targets and goals. There are no safe recipes for countries to reach these goals and targets. VET policy is a matter of chaos and complexity. Each country has to examine its own system. Examples from other countries can be used as good practice, but should be adapted to one's own national situation (e.g. Resnick and Wirt, 1996). One important lesson for systems

innovation is that the learning and experimentation should be coherent and compatible at all systems levels; from law, through institutions and organisations, up to teaching-learning processes.

The basic perspective in this chapter is the idea that changing complex social systems need coherent, persistent and consistent political actions at all levels of the systems, e.g. political debate and legislation, institutional arrangement, organisational and financial conditions and that the design of skilling systems and processes should be in line with each other. To realise system flexibility, change is needed at all system levels. Evaluation of systems change is complex in itself (Sanderson, 2000). Evaluation is balancing between inside and outside; evaluators are coactors in the system. Evaluation is organising critical reflection and learning moments. Evaluation is 'abductive' (Levin-Rozalis, 2000): n = single case-studies are interpreted in existing knowledge schemes through rational reasoning. Complexity of systems change makes simple/single evaluation useless and impossible. Goal-directed evaluation of changing systems is contradictory to the multi-actor aspect, connected to different stakes of different actor groups. Sanderson argues for evaluation as policy learning, where evaluation creates a relatively depoliticised setting for a rational discourse.

Three perspectives can therefore be discerned for discussion in the evaluation of VET systems change:

- (a) VET approach: flexibility and attractiveness; goal oriented versus development oriented; combination of economic and social goals; instructive versus constructive methods;
- (b) systems innovation approach: embeddedness, consistency between systems layers, interactivity; prescriptive versus facilitative; change concepts (incremental, dissipative; evolutionary), sources and models (used reference practices, e.g. other EU-systems);
- (c) evaluation policy: evaluation as policy learning versus evaluation as top-down (educational) science as policy reference source.

These three perspectives deliver three sets of descriptors to describe systems change on national VET-systems. This forms a frame of analysis to describe two cases of systems change: Denmark and the Netherlands.

2. Relationships between research, policy and practice

Before we perform an in-depth evaluation of educational systems change policies we should explore the relationship between policy and practice on the one hand and evaluation and research on the other. In a study by Caplan (1979), three theories are outlined to explain possible gaps between research and policy utilisation:

- (a) 'policy constraint theory' – policy systems and policy-makers are unable to handle the findings of educational research;
- (b) 'knowledge specific theory' – research has a too limited foundation and is located within a too narrow a framework of theories;
- (c) 'two communities theory' – policy-makers and researchers adhere to different cultures, each with its own norms values and agendas that are not explicit.

There are models that try to describe the knowledge generation and knowledge utilisation processes in educational research:

- (a) 'problem solving model': problem definition is followed by the identification of gaps in the knowledge area that leads to the identification of a study. Once data are made available, these are interpreted in terms of its policy implications which may then feed into the policy process. In this model there is no direct linkage between the different institutional actors;
- (b) 'interactive model': the use of evaluation research is part of a broader process of knowledge generation for change. In that process there is a complex web of relations between the different institutional actors at all levels. By stimulating interactions between key actors at strategic moments it is considered possible to strengthen the role of evaluations in the decision-making process;
- (c) 'political model': finally, this third group of models denies any direct rational role of educational research in policy-making. Here, results from evaluations and other research knowledge only come into play when they can be used as an argument to legitimise or even postpone policy decisions (Weiss, 1979).

In the OECD study *Educational research and development* (CERI, 1995) it is argued that a linear

model of research utilisation is to be avoided, concluding that research should be valued as much for its illuminative role as for any direct 'results' it might produce. Sutton (1999) provides a broader analytical framework of the policy process, which might be used to understand the design and potential usage of evaluations in a reform and a broader policy formulation process. She presents different models which do not see policy-making as a rational linear process.

Those models are:

- (a) the 'incrementalist model' that looks at a small number of alternatives for dealing with a problem and tends to choose options that differ only marginally from existing policy;
- (b) the 'mixed scanning model' which covers the middle ground between a rationalist and an incrementalist model. It suggests taking a broad view of options and looking further into those that require a more in-depth examination;
- (c) 'policy as arguments', a process in which policy reforms are presented as reasoned arguments. Policy is developed through debate between institutional actors. Language shapes the issues at hand in these debates, moulding social reality according to outlook and ideology;
- (d) 'policy as a social experiment'. This sees social change as a process of trial and error (experimentation) which involves successive hypotheses being tested in contexts in an experimental manner. It has its origin in the experimental approach of the natural sciences;
- (e) 'policy as interactive learning'. This approach is rooted in a criticism of development policy as being top-down and not generated from the communities in which policies are implemented. It argues for an actor's perspective emphasising the need to take into account the opinions of different institutional actors that have a stake in how a system evolves. The approach promotes an interaction and sharing of ideas between those who make policy and those who are most directly influenced by policy.

Table 1: **Basic beliefs of inquiry paradigms**

Item	Positivism	Postpositivism	Critical theory	Constructivism	Participatory
Epistemology	Dualist/objectivist: findings true	Modified dualist/objectivist: findings probably true	Transactional/subjectivist: value mediated findings	Transactional/subjectivist: created findings	Critical subjectivity in participatory transaction, extended epistemology of an experiential, propositional and practical knowing, joint findings
Nature of knowledge	Verified hypothesis established as facts and laws	Non-falsified hypothesis that are probable fact and laws	Structural historical insights	Individual reconstruction coalescing around consensus	In communities of inquiry embedded in communities of practice
Methodology	Experimental, manipulative, verification of hypotheses, chiefly quantitative methods	Modified experimental/ manipulative; critical multiplism, falsification of hypothesis, may include qualitative methods	Dialogic/dialectic	Hermeneutic/ dialectic	Political participation in collaborative action; primacy of the practical; use of language grounded in shared experiential context
Quality criteria	Conventional benchmarks of 'rigor' internal and external validity, reliability and objectivity	The same as for positivism	Historical situatedness; erosion of ignorance and misapprehensions; action stimulus	Trustworthiness and authenticity	Congruence of experiential, presentational, propositional and practical knowing-leads to action to transform
Inquirer posture	'disinterested scientist' as informer of decision makers and change agents	The same as positivist	'transformative intellectual' as advocate and activist	'passionate participant' as facilitator of multivoice reconstruction	Primary voice manifest by self-reflective action; secondary voices in illuminating theory, narrative movement and other presentational forms
Action	Not the responsibility of the researcher; viewed as advocacy or subjectivity; and therefore a threat to validity	The same as positivism	Found especially in the form of empowerment; emancipation anticipated and hoped for, social transformation	Intertwined with validity, inquiry often incomplete without action on the part of participants	The same as constructivism
Control	Resides solely in researcher	The same as positivism	Often resides in 'transformative intellectual'. In new constructions, control returns to community	Shared between inquirer and participants	Shared to varying degrees
Source: Lincoln and Guba, 2000					

Parallel to the debates on the nature of policy processes, Guba and Lincoln (1989) provided a coherent voice of criticism concerning the usage of evaluation methods that understood the task as a rational, neutral task to uncover 'the truth'. It was necessary instead to understand the complexity and the contextual nature of the evaluation process in what they defined as 'fourth generation evalua-

tion'. The rationales behind 'fourth generation evaluation' and the underlying methodologies will not be discussed further in this paper. However, a paper from 2000 (Lincoln and Guba) presents a useful overview on different paradigms of evaluation that lay out the purpose of the evaluation in the policy process and also determine the role of the different institutional actors in the reform process (Table 1).

3. Educational policy research: setting the stage

Educational research including programme and policy evaluation is expected to contribute to a deeper knowledge and insight on which policy-makers can draw, and act. While educational research is seen as contributing to highlighting problems and challenges in VET, there is an increasing debate about the degree to which and the circumstances under which a strengthened relationship may be formed between the three communities; policy-makers, researchers and practitioners. The question is if it could lead to stronger knowledge generation in action for practitioners faced with concrete problems, as well as to strengthened theory/practice knowledge base for educational policy leading to learning-based innovations.

The overall argument for a change in evaluation practices for sustainable reforms could be based on the hypothesis that an acceleration of knowledge construction and destruction is taking place, as brought forward by Lundvall (2000). Therefore, the argument goes, the issue is not so much having access to knowledge *per se*. The key to success is learning (and forgetting). This calls for more integrative relations and processes at different institutional levels, a complex that is traditionally characterised by narrow specialisations in terms of knowledge field and operational domain between the different institutional actors. Set in an evaluation context this also means a change in the purposes and framework of evaluations linked to educational reforms. Through narratives and dialogues policy evaluations may become an instrument of critical reflection and reinvention involving all the key actors rather than primarily being a way of measuring outputs and performance. This understanding of evaluation as a learning process will form the framework for this chapter.

In the early 1990s the Ministers for Education had already addressed the topic from a somewhat similar perspective in the following *communiqué*: 'In general, the level of investment in research and development in education and training is far lower than in any comparable size. The potential of educational research as an integral element of

improvement remains largely underdeveloped whether at the national, regional or local level. More important still, much research needs to be grounded in practice involving staff and institutions whether individually or collectively. To achieve this, innovation and experimentation should be given strong support' (OECD, 1995).

The British white paper on *Modernising government* identified learning from experience as a key principle in policy-making. Specifically the white paper stated: 'Government should regard policy as a continuous, learning process not as a series of on-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 more use of pilot schemes to encourage innovations and test whether they work.'

The growing policy focus on educational research and evaluations during the 1990s has been driven by different motives. They span from concerns about accountability, to a need to develop new systems of VET governance based on framework regulations rather than rules. Overall there has been a growing interest in understanding the driving forces behind systemic change and how innovation in VET can be spurred through educational research in economies where learning and knowledge are generally agreed to be of increasing importance (Foray, 1997).

Whereas evaluations in VET have traditionally focused on outcomes and accountability, there is increased interest from governments in new evaluation paradigms that can spur innovation and improve public policy through organisational learning processes. Argyris (1992) defines organisational learning as a process of change in organisational structures, codes or practices, that is triggered or reinforced by new experiences, new interactions or new information. Hart provides a useful analysis of the different ways in which organisations 'learn' about the effectiveness of their policies and practices (Kelleher and Hughes, 2000), as Kelleher has shown in an article on VET as a learning organisation (Kelleher, 2001). He

links these to different functions of evaluation within an organisational context.

- (a) The evaluation can be seen as a 'feedback mechanism' on the effectiveness and efficiency of policy (and practice). This view is grounded in cybernetic control and communication theory (Argyris and Schön, 1978).
- (b) Evaluation can be a 'stimulus configuration' for the creative minds of policy-makers (and practitioners) which is essentially a psychological perspective based on cognitive development models (Neisser, 1976).
- (b) The evaluation process can be a 'communicative domain' in which debate, persuasiveness of argument, weight of opinion and conflict of attitude and value systems will all contribute to a social construction of policy theories, ensuing policies and thus practice.

Based on two Dutch cases, Van der Meer expands the issue of learning through evaluations further by posing the following questions in relation to the evaluation process (Van der Meer, 1999):

- (a) when and how are evaluations identified?
- (b) when and how are they read or listened to?
- (c) how are they interpreted?
- (d) how do they become related to organisational characteristics or behaviour?
- (e) how do they eventually lead to changes in structures or behaviour?

Van der Meer concludes that if evaluation is to enhance organisational learning the following should be considered:

- (a) impacts of evaluations are constructed by actors, using their existing repertoires (defined as stabilised ways of thinking and acting among one or more of the involved actors);
- (b) construction processes do not take place in isolation but in the context of other actors and other construction processes, which supports our argument that reforms of VET should be understood in a broader institutional context;
- (c) a decisive point in determining the substantial impact of evaluation lies in the linking of new elements in the evaluation to elements in existing repertoires and/or to their contribution to loosening existing anchors of repertoire elements;
- (d) the mode and content of evaluation should link to the repertoire of one or more of the involved actors. Knowledge of repertoires can help to initiate explicit discussions on central presuppositions and value dimensions embedded in these repertoires, which may enhance interaction processes and linkages;
- (e) knowledge of different repertoires may help tailor evaluations to trigger unconventional interactions in the policy area which may give rise to new types of learning processes. An essential element of evaluating, aiming at learning, should be to take account of the contexts as perceived by the actors in the policy arena.

4. An analytical framework for case studies

The notions from the former chapters are used to develop an analytical frame to describe the use of evaluation in VET system change policies in Denmark and the Netherlands. Three main clusters of items are in this frame: the VET-approach; the systems change approach and the evaluation policy. The emerging knowledge economy causes paradigm shifts in the way people are becoming and remaining skilled for work. This shift demands changes in the institutional arrangements around employment and especially around VET, which makes the governance of systems change in VET a complex and subtle enterprise. Evaluation experts argue that a learning policy is the only way to deal with this uncertain and complex process; a top-down approach is a dead-end road in complex systems change.

Governance of change should be designed as a learning and experimenting route in which interactivity and creativity are the basic ingredients and in which the regulatory load is as lean as possible.

The case descriptions are quite process-oriented: the process of policy-making and the use of research and evaluation are tools in this policy-making process. The descriptions should give a glimpse of educational philosophies but this article is not primarily about the 'outcomes' of VET policies in terms of system characteristics.

After the case descriptions, a meta-analysis is done to answer the question of how evaluation is used in both systems change processes, ending in some conclusions and developmental views (a time line evolutionary perspective).

5. The Danish case: VET Reform 2000

This section introduces the most recent reform of the Danish VET system – Reform 2000 – with a discussion of some of the main policy and pedagogical challenges that formed the basis for the reform. Both from a policy and from a pedagogical perspective the Danish reform was initiated, designed and implemented in ways that consciously aimed at fostering a learning environment in the interplay between policy formulation and the invocation of what, from the outset, was defined as a new ‘learning topography’ (Christensen, 1999).

5.1. Policy design in a climate of change

The Danish VET system is deeply rooted in the tripartite system. Prior to the reform this was for

example, reflected in an entry construction for each industrial branch and sector mirroring a traditional industrial structure. There were intense debates on how VET within the tripartite structure could deal with a profound reinvention of its systems and institutions to meet changing system requirements. One of the basic arguments was that an accelerated pace of innovation change calls for changes in the whole curriculum and didactic thinking as depicted in Table 2.

A trend in the learning economy is that innovation cycles are getting shorter (IPTS, 2000), and constantly being disrupted by technological innovations in a creative destruction process, as Schumpeter first pointed out. It could be argued that this is one of the driving economic forces behind an increased policy focus on lifelong learning and emphasis on core skills. When existing knowledge repositories are more quickly

Table 2: **Setting a new context**

Issue	Industrial economy	Learning economy
Markets	Stable	Dynamic
Scope of competition	National	Global
Organisational form	Hierarchical	Networked
Organisation of production	Mass production	Flexible production
Key drivers of growth	Capital/labour	Knowledge/innovation
Key technology driver	Mechanisation	Digitalisation
Source of competitive advantage	Economies of scale	Time to market, innovation
Relations with other firms	Single mover	Alliances and collaboration
Workforce		
Policy goal	Full employment	Employability
Occupational profile	Fixed tripartite regulated professional identity	Converging and continuously emerging and being reshaped; tied to contexts and technological advance
Skills	Job specific	Multidimensional (deep and broad foundation skills)
Requisite education	A skill – a degree	Lifelong learning
Source: Adapted from Atkinson, 1998.		

made obsolete, the demand for 'learning to learn' skills and other core skills grows in importance as a central educational concern.

In a Danish reform context this translated into demands for a legislative framework that could accommodate continuous change in the curricular structure and supply as a result of changing skills demands, without necessitating a change in the legislative base. The institutional settings in VET have traditionally focused on skills formation through instruction and imitation. With the reform, the learning environments would change from instruction and teaching to a situation where the learning agent is more successful in pursuing his/her own goals, fostering learning-to-learn competences.

5.2. Preparatory phase

From the end of 1997 and through half of 1998 there were extensive discussions within the tripartite system and preparatory work was carried out which formed the basis for legislation 234, the so-called Reform 2000. The legislation was approved by the parliament in April 1999.

In the preparatory work, the Ministry looked to international research and experiences. From Finland, the modularised structure with individual pathways has influenced current thinking about flexibility in the organisational structure and supply. The concept of an educational topology introduced by OECD (1998) has been influential in the new terminology. In this landscape the student is an active voyager and discoverer, defining his or her path rather than simply succumbing to the mode of instruction and the predefined and fixed curriculum fabric of the educators.

An internal task force was constituted to examine how these concepts could be translated into a new legislative practice and to critically examine current barriers in the legislative frameworks. On this basis they were to provide some governing principles as a foundation for a new model for a descriptive curricular framework (Danish Ministry of Education, 1998). The new descriptive framework should enhance three central areas:

- (a) individualisation of progress, content and form;
- (b) more emphasis on learning rather than teaching;
- (c) enhanced and more flexible ways in the descriptive framework; learning offered in situational arrangements to allow for a holistic approach; integrating technical, practical, personal and general competences.

The main recommendations from the group were:

- (a) more flexibility in the descriptive frameworks and taxonomy to ensure dynamics and continuous innovation in the curricular offer;
- (b) a clearer organisation of the descriptive levels, including an introduction of a new element of authorised administrative teaching guides to ensure a continuous adaptation without necessarily changing the legislative frameworks;
- (c) replacing traditional taxonomies to provide a uniform terminology in the description of goals and ways of communicating intentions at the different levels;
- (d) quality development should be constituted at a central legislative level so that quality intentions, indicators, criteria and verifications can be defined in the individual course programmes and educational guides.

The recommendations from the task force prior to the reform later set the stage for the follow up and evaluation of the first piloting of the reform (Shapiro et al., 2000).

5.3. Creation of an environment of change

Parallel to the introduction of the reform, the Ministry took several steps to create an environment of change in the implementation of the reform. The most important actions were:

- (a) a study for a new didactical paradigm for the reform, providing a model for a descriptive framework in line with the intentions of the legislation. This would replace existing taxonomical thinking and at the same time govern the description of goals and pedagogical arrangements (Shapiro and Christensen, 1999);
- (b) from summer 1999 up to the full implementation in January 2001, the reform was gradu-

- ally piloted in several schools to better understand at an early stage structural, organisational and pedagogical implications which might require further policy actions;
- (c) the Ministry initiated several training of trainers initiatives with the support of the Danish vocational educational teachers' organisation;
 - (d) several development projects were commissioned to practitioners in the schools with process support from DEL (Danish Institute for educational training of vocational teachers) concerning the development of tools and practices that could support the change processes ⁽¹⁾;
 - (e) the Ministry commissioned the specification and the development of a web-based infrastructure and tool, 'Elevplan'. Its goal was to enable the schools to create an environment that could manage modularisation and support students in the individualisation of choices and planning of their education. The specification of the web-based tool was developed with the heavy involvement of school practitioners and other actors with expertise in the reform;
 - (f) two concurrent evaluations in the pilot phase of the reform which will be described in the following sections.

5.4. Main new elements in the reform

Reforming the descriptive frameworks involved both a continuation of known elements and development and integration of innovative elements. Among the most important known elements were the general ministerial order, the education order and the local educational plan.

The most important innovative elements were:

- (a) the student's individualised and documented learning pathway;
- (b) a logbook, possibly developing as a learning portfolio;
- (c) guidelines for the ministerial orders;
- (d) an expansion of the ministerial order with

provision for evaluation of results and quality development in connection with the educational offer;

- (e) a modular structure in the curriculum combined with a broader descriptive framework.

The legislation was amended in order to emphasise and legitimise the individual student's opportunities as the active party in the learning process, shaping his or her individual learning process. The legislation operates with a general and an individual axis. Central, potentially innovative instruments were legitimised through the main legislation:

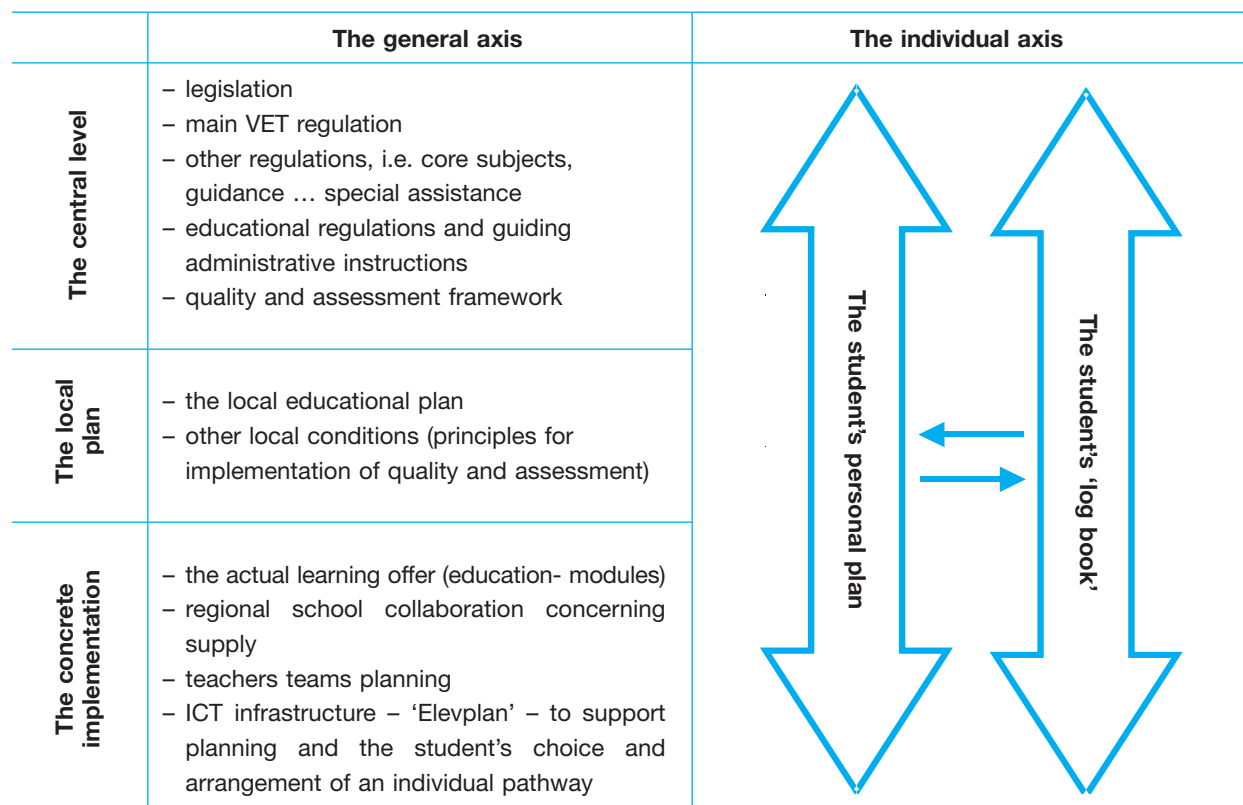
- (a) the student's personal educational plan, laying out the roads and the pathways in the learning landscape;
- (b) the logbook, as a practice narrative and a linkage between school and workplace learning in the firms;
- (c) the contact teacher, facilitating the development of the student as a didactician and in ensuring the linkage between the collective and the individual pathway.

The reform and the didactic framework developed (Shapiro and Christensen, 1999) were to offer a pragmatic basis and a terminology that provided enough space for concrete and local imaginative and reflective actualisation. Not as a well-defined master plan, one model suits all, but as a basis for planning situational learning arrangements in a Danish context defined as guided participation (Nielsen and Kvale, 1999).

The student should therefore become an active agent in gradually learning to define and reflect on his/her own way through the learning landscape rather than merely following an externally predefined signposted road. The metaphor 'the student as a didactician' is used in the pedagogical paradigm to denote this change. It means that the student gradually engages in a reflective planning process and takes possession of the didactical space. This arena has traditionally only been inaugurated for actions by planners whose role was to lay the choreography and by the teachers whose role was to direct and instruct through the lesson plan (Shapiro, 2002).

⁽¹⁾ Many of these projects have later been published on the web so that they are widely accessible.

Figure 2: The legislative framework



5.5. The evaluation of Reform 2000 in Denmark

5.5.1. First evaluation: November 1999 to February 2000

Most schools were already collaborating within a regional structure prior to the reform. In the first pilot phase starting in August 1999, five of those regions participated in the pilot scheme with all or some of the new ways into the entry programme. In early October 1999, the Ministry of Education sent out a restricted tender to carry out an evaluation of the first experiences with the reform. The Ministry of Education consciously avoided the use of the term ‘evaluation’ calling it ‘a reflective gathering of experiences’. The aim of this first evaluation was to inspire and assist the Ministry, the directors of the schools, the teachers, the National Institution for Vocational Education and the social partners (the local councils) in the further implementation of the reform. The evaluation was to identify both current barriers and diffi-

culties and those promising new practices that could guide reform implementation or where further policy interventions were needed. After the tender brief, the Technological Institute was commissioned to carry out this evaluation, which took place between early November 1999 and early February 2000.

The evaluation included a survey among all the school directors involved in the pilot, at top and middle management levels, a sample of the teacher population and a sample of the student population. The latter was complementary to a case study.

The themes to be covered were:

- perceived changes in work and professional profile and how that was dealt with;
- knowledge about the reform and how the actual implementation was perceived among the different actors;
- information and collaboration among the different actors;
- implementation models, qualification demands, frameworks and conditions, the overall vision;

- (e) experiences with roles, instruments and structures;
- (f) the contact teacher, organisation and qualification for the new role.

5.5.1.1. *Approach taken*

As part of the design process, both surveys and the qualitative frameworks were presented to and discussed with representatives of all the key actors to ensure engagement and relevance of the process.

Parallel to the collection and analysis of survey data, focus group interviews were set up with members from school councils and teachers' unions. The 29 focus group interviews were supposed to target coordinating forums at national regional and entry levels. During the evaluation it transpired that many of those forums had not yet met and that the evaluation was instrumental in putting 'the breath of life' into these areas.

Despite the tight time scale and implementation pressure in many of the schools during the first phase, our experience was that the participatory evaluation design resulted in open and willing desire to share and debate the implementation process. A common comment from many of the participants in the evaluation exercise was that the focus group meetings led to an important dialogue between the participants that helped shape some of the implementation efforts.

5.5.1.2. *First findings*

In one region, a centralised approach was taken to the first implementation phase. The centralised approach was a conscious decision from the board of directors in the school region to ensure consensus in the overall implementation approach. The belief that it could be used to advantage in speeding up the implementation process by facilitating the sharing of developed materials and tools among the schools concerned, was in fact borne out. Judged from that perspective the approach was quite successful and had the advantage of developing a strong collaborative network among middle management. The teachers, however, were critical of this approach as in their view their involvement in the decision making process came too late. Major decisions had already been taken, having a demotivating effect and giving rise to

organisational difficulties in terms of implementing those decisions.

In the other regions a more decentralised approach was taken where learning by doing was the central development strategy. This ensured that many of the teachers felt more motivated to work the reform in practice but it also led to considerable delays in developing central pedagogical tools. This approach made it difficult to balance an ideal pedagogical development process and the daily operational practice with the concrete demands from students and other stakeholders.

In the second pilot phase the number of schools and means of entry were extended. There was more focus on the middle manager's role as a catalyst of change not only in relation to changing pedagogical process and practices, but also in relation to their role as carriers of debate and information. As the reform process progressed, the National Institute for Vocational Education developed different training programmes and approaches for middle management acting as reform consultants. Directors from schools, the teachers and their representatives have pointed to this programme as a central component in the implementation process. From August 2002 a comprehensive competence development programme organised as action learning has been implemented in collaboration with the Ministry of Education and the National Institute for Vocational Education where the local reform agents play a central role.

Unsurprisingly, both managers and teachers found that the reform increased their workload considerably in the first phases. According to surveys one of the biggest changes for both managers and teachers is the reshaping of professional identities and roles. Management and middle management had to work more proactively with information and delegation strategies, and they have had to develop a planning understanding with a much tighter fit between leadership, administrative planning, logistic and pedagogy. More importantly, they have had to learn to be carriers of constant and value-based change. The reform agent training initiative has been central to bringing about this change, as have the meeting forums established between the schools. Many teachers have had to identify themselves as facilitators rather than

teachers in the traditional sense having to begin working in communities with a broader branch and technological perspective. For many this was and is still a difficult process and has in many cases led to considerable stress.

After the first three months 25% of the teachers surveyed considered looking for another job. One solution already implemented is the formation of teachers' teams with different occupational profiles. The local development projects started with financial support from the Ministry have also been vital in introducing new professional identities for those teachers directly involved in developing new tools and materials.

5.5.1.3. *The didactical framework*

Prior to the reform many of the schools had already been engaged in a more project and problem based learning environment. An initial assessment of some of those projects, however, showed that much of the so-called project and problem-based learning was based on a pre-defined problem area.

Small schools had particular problems with the combination of cross-curricular integration and individualisation at the same time. The student base was simply not big enough to offer a variation in projects that could also be chosen at different levels and/or learning styles, even when the traditional classroom structure was abolished. One of the central questions concerning individualisation which is still on the agenda is how to design those learning frameworks so that they accommodate the needs of different student groupings, not only in an academic and professional sense, but also socially. Some schools consciously chose to have a more limited selection of project opportunities so that students would not constantly change between teachers and different groupings of fellow students.

'People need both a social and a professional reference framework to develop. In the constant hunt for an individualised pathway we might easily neglect the students' personal needs and reference points which in fact constitute their learning and development basis' (Shapiro et al., 2000).

5.5.1.4. *The new student role and the contact teacher*

Both the survey and interviews conducted showed a high degree of student satisfaction. In

the survey 67% of the students responded that they were always, or almost always, thriving in their school environments. The first evaluation showed that older students, regardless of their background, thrived in the new environment with a higher degree of autonomy and possibilities to plan and influence their educational pathway.

A central new element in the reform has been the definition of the contact teacher function. For the students the contact teacher is an important anchor in developing their 'learning to learn' and 'learning to plan' qualifications, even if at the time of the first evaluation several of the tools to support those processes were not in place. Many contact teachers also felt that this role enabled them to give more varied support strategies to different student groups.

5.5.1.5. *Dissemination of findings*

The results of the evaluation were published in a main report as well as on the web, comprising findings and a write-up of main findings plus an annexed write-up of all the interviews and survey data. A one-day seminar was organised with several hundred participants, where the findings were presented and debated in a series of workshops. Regional workshops were also held with their point of departure in the evaluation findings. This formed a central element not only in exchange of experience but also in the creation of a 'common reform narrative'.

5.5.2. **Second evaluation: October 2000 to February 2001**

From January 2000 and again from August 2000, the pilot was expanded so that it involved almost all schools and entries that were approved by January 2001. Where the first pilot phase comprised only the programme basics, the pilots from 2000 also included the specialisations within some of the programmes.

In early September 2000, the Ministry of Education sent out a second restricted tender. The evaluation took place from mid October 2000 until the end of February 2001, carried out by the Technological Institute. Again the aim was to draw attention to barriers and problems in the implementation of the reform as well as practical solutions and strategies undertaken to realise the intention and aspirations of Reform 2000. The evaluation was to include the eight regions and all the entries. Additionally, the evaluation was to

comprise those schools that were piloting the specialisations at this stage (storage and logistics, train assistance, forestry, chauffeur). Survey themes were mostly built on the first survey and able to capture changes in perceptions and practice over time. A sample survey was performed among students that had dropped out of the programme/entry they had originally chosen (telephone interview sample of 410).

A qualitative analysis comprised 16 focus group interviews with managers and teachers, 14 interviews with apprenticeship companies at both apprentice and employer levels and 56 interviews with local educational councils and teacher union representatives.

Themes covered were:

- (a) organisational structures and changes;
- (b) the pedagogical space;
- (c) instruments and roles;
- (d) qualifications measures and organisational development;
- (e) experiences with roles, instruments and structures;
- (f) information and collaboration;
- (g) companies and the apprenticeship scheme.

5.5.2.1. *School organisational initiatives*

Whereas the first evaluation showed that the organisation of the basic programme in all school regions built on the same five-week module structure and with a fairly fixed duration of 20 weeks, considerable change had taken place in the following period, also tied to more variations in pedagogical models. The second analysis of experiences showed many different models for organising the curricular structuring. Several schools were working towards models more flexible than the five-week module. This also involved innovation in the layout of the physical settings and organisational frameworks, using for example, 'own learning time' where students work individually or in groups on a project or a particular task with access to some counselling by appointment.

In response to converging sectors and branches and to meet requirements for more flexible qualification structures, one innovation in the reform is that the many individual trades and professions are now organised in six broad entries in the basic period:

- (a) technology and communication;
- (b) building and construction;

- (c) crafts and technique;
- (d) from earth to table;
- (e) mechanics, transportation and logistics;
- (f) service.

This has also led to demands for a more holistic and integrated approach to curriculum development across topics and trades within the specific entries. In the first round, responses to these challenges had at times attempted some integration. Students with an apprenticeship for a particular profession would, for example, switch to a school where particular facilities were in place for a particular trade. There were also examples of team-teaching in integrated workshop settings where facilities permitted. At the time of the second evaluation there were still unsolved problems concerning family formation between similar trades versus a wider integration across all trades in entries. Problems such as teacher competences, the physical setting, equipment, as well as 'VET-school traditions' remained. The first evaluation showed that different groups of schools tried to compensate for some of these problems by giving the students the opportunity to change schools. In the second round students who had chosen a particular trade were well aware which schools covered which education opportunities and chose accordingly. This development is ascribed to the fact that the initial counselling and guidance of students at the time of the second evaluation was now almost fully implemented in the schools.

The second evaluation showed more flexibility in the duration of modules and basic programmes and of content and learning style. Parallel to the first pilots, the Ministry of Education together with teachers and other specialists had initiated a specification and development of a web-based tool to facilitate modularisation, planning and selection of activities for both teachers and students. In the second evaluation the implementation of the tool began with the positive expectation from both teachers and administrators that it was not just another administrative device, but that it could enhance the delegation of planning autonomy to teachers groups and facilitate students' planning competences.

5.5.2.2. *Variations in strategies with different student groups*

The different groups of schools were now actively working with different strategies to make

allowances for students who had not made their final education choice and for the educationally or socially weak students. This occurred through various pedagogical initiatives and through different forms of flexibility during the basic programmes. The flexible length of the basic programme created problems for the educationally 'strong' students taking supplementary courses thereby prolonging the basic programme prior to the apprentice scheme, because apprentice companies still had a notion about a fixed duration programme. One of the pillars of the reform has been the concept of an individual pathway for a student to develop their planning and learning to learn competences. In reality some of the weaker students are neither ready nor motivated to engage in this co-planning and continuous selection process. They tend to select and plan as their closest friends. Since the evaluation many schools have started developing the planning and selection process as a continuous and competence-development process, rather than regarding it as a competence all students have when they start. There has also been debate on whether students who had already made their choice should have this challenged, or whether the 20-week basic apprentice pathway should be shortened.

5.5.2.3. *Project pedagogy and assessment*

Whereas the organisation of subjects in cross-curricular projects is perceived as positive by teachers and students alike; selection and planning was seen as too complex. Some schools have therefore chosen to offer shorter projects or projects with a flexible duration. Students thus have more options to choose from across modularised subjects in families rather than across the whole entry which may consist of as many as 17 different trades (such as in the entry Crafts and Technique).

Overall, programme projects began to have a more prominent role than previously with the integration of theoretical subjects such as IT and mathematics, combined with more focus on new, more process-oriented assessment methods that can support the project work method. However, at the main programme level the ministerial order on basic subjects was seen as a barrier to new forms of assessment. After the second evaluation was carried out, the Ministry of Education has

initiated development work in collaboration with schools to look at alternatives in assessment.

5.5.2.4. *Teachers' involvement in the change processes*

As a consequence of the reform many schools have delegated more autonomy to the teachers, giving them more direct influence and responsibility for administrative and pedagogical planning processes. The second evaluation showed, however, that teachers continue to have little influence on conditions relating to financial management, which traditionally is the responsibility of the management. A typical explanation from middle management would be that finances were so limited that they were afraid to relinquish control of the budgets.

Prior to the planning of the common action learning programme, some schools voiced a need for acquiring new skills, not only relating to changing pedagogical and organisational practices, but equally to acquire tools and methods. Consequently, pedagogy and financial and administrative management could be more tightly interwoven in the pedagogical practice.

5.5.2.5. *Instruments for pedagogical innovation*

Approximately two thirds of the teachers surveyed in the second evaluation have had the opportunity to participate in development or planning activities related to innovation in pedagogical practice and curriculum.

Parallel to the pilot of the reform, the Ministry of Education had a funding mechanism where groups of schools could apply for funds to initiate particular development activities of relevance to the reform. Those development projects, at times together with process consultancy offered by the National Institute for Teachers Vocational Education, have been a central source of qualification. An explanation is that participants have been able to relate them to practice and practical problems in an experimental and reflective manner, as such functioning as communities of practice. However data from both surveys and interviews have shown that the research and development projects outside their particular environment have not had as wide an organisational impact as was intended. One reason seems to be that organisational frameworks and communities were not created as part of a wider pilot organisation so that they could ensure the organisational

anchoring of the research and development projects' processes and results. Consequently, the research and development projects often ended as tacit knowledge among the participants and of little practical use for wider groups. Since the second evaluation, the Ministry of Education has decided to have all the research and development reform projects analysed and structured as a pedagogical resource base as part of the action learning scheme.

5.5.2.6. *Roles and governance structures*

In many ways, the organisation of the wider entries mirrors developments in the labour market concerning convergence of trades and jobs. Within tripartite educational systems such as the Danish, the evaluation data raises the question of how these developments can best be reflected in the organisation of trade committees in the future. This is not least an issue when it comes to emerging new job profiles for which there is not a national trade committee or union. Evaluation data highlight the issue of whether the traditional tripartite structure is the most appropriate to develop a more dynamic and concurrent flexibility in VET programmes, for example in technology and communication that have undergone heavy economic and technological fluctuations in an economic down time. In a wider perspective the risk may be that vocational educational programmes will be outdated before they are even implemented and therefore replaced by certifications with a much shorter employability perspective.

5.5.2.7. *The pedagogical space*

Students' joint responsibility for their own learning has given teachers many challenges, since few students have been able to define their own learning pathways and successively be confronted with different choices of a more or less complex nature. For educationally and socially weak students, some teachers have worked with planning and choosing skills that progressively should be developed as lifelong learning. For some students the concept of joint responsibility for their own learning has in their view turned into a laissez-faire attitude from teachers not sufficiently challenging their choices.

5.5.2.8. *Roles and instruments*

The professional identity of the teacher as an expert of a certain trade is profoundly challenged

in favour of a role as a process facilitator being able to handle professional and pedagogical breadth and depth with both students and in the team of colleagues. Establishing open and more problem-based and experimental learning settings and the higher degree of individual guidance of students tend to provoke many teachers' basic understanding of themselves as the professional master of the learning space. The uncertainty as to whether or not the educational goals can be reached through their teaching makes teachers ask for particular tools to understand the projects' subject-related complexity and to link them safely and uniformly to the subject-related standards. This concerns both the teacher's perception of what constitutes and should constitute their professional competence and identity, but it is also related to a concern about 'safeguarding' the students and giving them what they perceive as proper guidance.

At the time of the second evaluation most managers felt extremely pressured by the financial frameworks. Data clearly showed that when managers felt uncertain about their own competence base they seemed to stick to old managerial routines, as inappropriate as they might be.

5.5.2.9. *The contact teacher*

The first evaluation already pointed to the implementation of the contact teacher as an immediate success in the eyes of teachers, management and students alike. This was also the case in the second evaluation. Among the many new instruments introduced with the reform, the contact teacher role is one area where the teachers feel they are best equipped to handle the challenges of the reform. In interviews with contact teachers they often found it difficult to draw the line between the educational tasks belonging to the contact teacher and those belonging to the student counsellor and ultimately to social agencies as more and more students are referred from social authorities with a variety of difficulties. It is a common perception among contact teachers that guidance of the academically weak may prevent dropouts and enhance their satisfaction with the school because initiatives such as special pedagogical support and special teaching to support the student's learning are introduced early in the programme. Interviews and survey data from students seem to support this view.

5.5.2.10. *The logbook and personal educational plan*

Survey data as well as interviews show that neither managers nor teachers believe that the personal education plan and logbook were useful for learning and the apprenticeship scheme. As in the first collection of experiences, the schools struggled to make the logbook an active instrument in the planning of the students' education and not a 'dead' file that is only updated on the initiative of the teachers/contact teacher.

Ideally, the logbook must also accommodate the variations in students' learning and development profiles in line with some of Gardener's theories about the seven intelligences (Gardner, 1997). Reasons for failure may be that much dialogue with the contact teacher takes place in specially organised sessions outside the concrete practice context and that the assigned contact teacher will have little previous knowledge about the student from actual practice in a teacher-student relationship. Furthermore, if we assume that 'know-who' and 'know-how' competences are contextualised and tacit in nature as argued by Nonaka and Lundvall (Nonaka, 1995), then it can be a problem for the students to reflect over those type of competences outside the actual practice.

5.5.2.11. *Apprentice companies' knowledge of and experience with the VET reform*

The second evaluation provided a first and limited view for apprentice companies on the VET reform. Even now, much work remains to inform apprentice companies about VET reform and the increased flexibility it offers compared with earlier arrangements; the lack of apprentice companies being one indicator.

To create comprehensive and integrated VET programmes, efforts and examples are needed to illustrate changing relationships between the school and the company that can best accommodate genuine apprenticeship-based learning. Cooperation appears to have worked best where the logbook has linked the apprentice, the company and the contact teacher in a joint effort. An example is that the logbook supports the student's work experience history to the company and simplifies planning of the next apprentice period by both the student and the company. The student can then, once back at the school, convey processes and results. In those cases the logbook begins to function as a learning portfolio which

ensures and codifies continuity in the interaction between the different players. In other instances, the student was still more or less the primary carrier of the professional interaction. To many students there often appeared to be a big discrepancy between the reality of the school and that of the apprentice company. The primary focus of the student is on shaping a professional identity and the school environment is often perceived as prolonging high school culture. Even though only one of the case companies, together with the school and the student, had implemented full operational use of the logbook as a model of cognitive apprenticeship (Collins et al., 1989), several interviewed companies had developed frameworks for the apprentice. A common characteristic of those companies was a gradual shaping and integration of professional and social identity through a planned delegation of and complexity in tasks and assignments within a working team. The students would perceive such a practice as a situation 'under control' and as if they were full members of a working community, while still learning and developing. For some students this was contradictory to the 'as if' school reality, where the broader frameworks and opportunities were perceived as a lack of plan and direction and where they did not necessarily perceive themselves as 'co-planners' in the school environment. This again raises the question of understanding joint responsibility for own learning and how this competence can be developed by a dynamic interaction between the school and work, based on training in the dual system while not creating an understanding of a split universe for the student.

5.6. Final comments on the use of evaluation in Reform 2000

When the Danish VET reform was designed and implemented the Social Democrats were still heading the government. The Liberal Party has since come into government taking a more critical stance on the use of research expertise to govern policy formulation.

When the two evaluations of the Danish reform were designed it was intended to create an evaluation and framework that was both useful and usable, despite severe time and budget constraints.

The evaluation framework was designed to:

- (a) understand the context of change: collect baseline information about the school environment, teachers and managers;
- (b) inform on what was possible, taking into account constraints tied to, for example, building infrastructure, variations in student populations across schools, etc., to record 'good practice' across a variety of realities and identify potential barriers that could be improved through policy action;
- (c) learn how change was organised and introduced: track processes and decisions and how they were perceived by different actors to identify successful models or components.

The likelihood of evaluation findings being used to inform and influence policy and practice will, to a large extent, depend on the form in which its findings are made available (Boaz, 2002).

Our experience with the Danish VET reform was, as Boaz also recorded, not only a matter of making available a range of written material, published both in print and on the web, to specific audiences but also our willingness and ability to present and debate findings and policy recommendations with different stakeholders. Often this is only budgeted for to a limited degree

and often does not take into account in the time scale laid out in the tender brief.

A reform of the commercial colleges has been approved and is currently under development. It can be seen as a continuation of Reform 2000 by expanding the structural flexibility of the system to provide for previous working experience, expanding academic qualification options within a vocational system and strengthening continuous entry/exit options by offering partially-recognised qualifications. These changes fall within the rhetoric of a lifelong learning agenda and improve system efficiency.

Evaluating findings clearly pointed to the need to tie skills development to daily working practice and to ensure a higher level of knowledge exchange between teachers across schools and regions. This has laid the foundation for developing a national scheme coordinated by the National Institute for Training of Vocational Teachers. However, the initiative builds on principles of action orientation and organisational learning with a strong involvement of local teachers and middle management acting as coaches and change agents. The initiative is very flexible in that different schools and school regions can tailor a programme both in terms of content and pedagogy, to suit their particular needs.

6. The Dutch case: implementation of WEB-1996

After 20 years of political debate and preliminary legislation, a new law on vocational and adult education was launched in the Netherlands in 1996. This law, the WEB, is built on several developments running since 1980. Nijhof (1997) describes this systems transformation in detail. The Dutch VET system changes are targeted at system efficiency therefore most transformation took place on the institutional and organisational levels.

Two major perspectives, regional and sectoral policies, played a role in the design of the WEB. The regional perspective has led to the formation of large regional training centres (*Regionaal opleidingscentrum* – ROCs) with a high degree of autonomy. The sectoral policy is based on the implementation of sectoral qualification structures, in which the labour-markets demands are translated into educational trajectories. These two perspectives are not completely complementary: with their qualification structure, social partners try to reduce the autonomy of the ROCs. The main sector in each region will be the critical factor for the implementation of the WEB. The Dutch parliament has obliged the Minister for Education to present a formative evaluation of WEB before the end of 2001. Consequently, in 2000, seven research teams have been working on several evaluation topics, which has led to an evaluation report in the summer of 2001 (Stuurgroep Evaluatie WEB, 2001). Some key results will be presented in the next section.

6.1. A short history of changes in Dutch VET

Until 1968, VET was closely regulated in the Netherlands. VET comprised a few small courses based on private initiative and not financed publicly. In 1968 a new law on secondary education was launched, in which vocational education was envisaged for everybody: after a period of general education each student should undertake vocational education in

preparation for the labour market. After each level of general education a form of vocational education was organised in the new system for secondary education. Lower or junior VET was for students directly from primary education; secondary VET for students from junior general education; higher VET or professional education and universities for students in general secondary education. Apprenticeship courses were not regulated in this educational law, but were provided for in separate legislation (De Bruijn, 1997).

Soon after establishing the 1968 law, the 'system' for VET was seen not to be as systematic and complete as intended. There was no uniformity in level, duration and design of courses and the structural supply did not cover all the educational demands. This was particularly so for graduates from the junior VET, with a shortfall in their training supply: labour-market demands were higher than their qualifications and the training and course supply within the educational system and the apprenticeship system failed to serve all these students. This became a major problem in the economic recession of the late 1970s and several experiments and initiatives grew to fill in this systemic gap.

Since 1969 two movements in the development of Dutch VET can be seen: a search for efficiency and a fight for accessibility of lower educated youth. The efficiency search began in courses for health care education. Steered by the Ministry of Education, 35 small courses were merged and organised in separate schools into a transparent system of care education. In 1983 this process ended with larger colleges for sociopedagogic education. In the next decade these colleges adopted courses for nurse education which, until then, had been in-service education under the responsibility of the hospitals. Around 1995 the merger between sociopedagogic education and nurse education was finalised. Alongside the process in the healthcare sector, the technical and economic educational infrastructure became involved in the merger policy: the SVM-law was launched in 1986,

targeting the merger of healthcare, economic and technical education into regional colleges for vocational education. 'S' stands for developing a strong VET sector materialised in the organisation of regional colleges. 'V' stands for educational innovation, of which the design of a qualification structure, supported by tripartite committees, was the most significant expression. 'M' stands for secondary VET. This merger process resulted in 45 regional colleges for VET around 1995. In addition, some specialised colleges and 12 vertically-organised agricultural colleges completed the educational infrastructure for VET in the late 1990s.

The fight for accessibility of VET for low educated youngsters took place alongside the merger processes. In 1969 the Minister for Education published a paper on the educational position of this group at risk. In the paper the rights of youngsters (under 18) to one day's schooling a week was recognised (although it had already been formalised in the labour law of 1919!) and translated in partially compulsory education (one day a week). Under the auspices of participatory education, a movement of educational innovation started to develop specific education for these groups. However, 10 years of political-educational argument still did not result in commonly accepted vocational paths. The economic recession on the end of the 1970s was a major cause of the failure: in 1980 the Minister for Education launched short full-time vocational education opportunities to fill the educational gap. From the employers' side these were viewed with suspicion, as they trusted only the apprenticeship model. In 1983, a committee for industrial development sounded a warning, stating that good vocational education for all should be the common responsibility not only of the educational system but also of the labour system. This had a shock effect, partly because it came from an unexpected source. All stakeholders agreed within a year on a unified system in which full-time and dual courses were accepted equally and in which the steering principle of a common qualification structure should be the main line of educational development. Each student should be qualified at an entrance level, comparable to the apprenticeship courses and both the labour and education worlds recognised the common responsibility for this target.

Consequently, the picture of VET in Holland at the beginning of the 1990s shows regional colleges in development and a mixed trajectory model combining dual tracks and school-based tracks for the lower and higher levels in secondary vocational education. The WEB, the 1996 law on vocational and adult education and training, forms the final legislative cornerstone. The regional colleges merge with the schools for the supportive training in the apprenticeship track and many kinds of adult education, ranging from second chance general education, through language courses for immigrants to in-house development of workers. The WEB is meant to round off the legislative efforts for building a responsive VET system, able to deliver all kinds of courses and training at secondary level.

6.2. Features of WEB-1996

The main features can summarised as follows:

- (a) integration of initial and continuous VET: under the WEB adult education (both vocational and general) and vocational education are combined in one coherent framework. ROCs should provide both types of education and training. The policy instruments and the institutional systems are unified;
- (b) social demands and economic demands: the main issue is preparing and supporting youngsters and adults for lifelong learning, according to the needs of a rapidly changing economy. Social demands are fulfilled by a general target for basic qualifications (qualified to participate in the economy). For all entry levels of students the system should deliver appropriate courses. Economic demands should be fulfilled through output definitions (steered by a qualification structure) and the opportunity for training centres to enter the training market. The output definitions should not be biased for training trajectories: it should be possible to reach the same output via different pathways; both company-based and school-based;
- (c) regional perspective: the training infrastructure is organised by region. The ROCs are mergers of the former sectoral vocational schools, the adult education centres and the apprenticeship support structure. Together

they form large community colleges with 15 000 to 30 000 students and apprentices: in fact they have become the largest educational organisations in the Netherlands. They cover all occupational sectors, except agriculture and some smaller sectors. The ROCs should cover all the occupational training needs on secondary level (EU level 1 to 4) in their region;

- (d) sectoral perspective: the needs of industrial branches and occupational sectors should be satisfied via the regulation of goal definition. A qualification structure contains all recognised courses, formulated in output terms. Social partners, organised in 21 sectoral bodies, are responsible for the definition of qualifications; the Minister for Education will legitimise all proposed qualifications, under the restriction of transparency. The ROCs have to apply for permission to deliver the appropriate courses for their region. Examination boards, in most cases connected to sectoral bodies, verify the quality of assessment and examination by the colleges. This enables, sectors and branches to have input to and a responsibility for the quality of the system;
- (e) autonomous colleges: within the boundaries of the qualification structure (product definition), the examination regulation (quality definition) and the application rules for courses (financial frame), the colleges are autonomous to deliver the courses. They can choose content, didactics and the organisation of practical learning to their own insights. They are responsible for a public quality report each second year and the inspection service of the Ministry checks their quality policies. The regional market should regulate their quality but their monopolistic position makes this inefficient.

To achieve this ambitious set of goals, WEB-1996 contains legal instruments both at institutional and organisational level. On the institutional level the main instruments are:

- (a) the qualification structure, in which a prescription can be found for 750 different educational paths towards occupations;
- (b) about 20 sectoral bodies, responsible for the definition of the qualification structure and the quality of learning-working environments in the enterprises; to realise this task, the sectoral

bodies organise the debate between social partners and educational representatives;

- (c) about 50 examination bodies, responsible for the quality assurance of examination procedures of the colleges;
- (d) the inspection service, executing the quality control for the Minister for Education.

At the organisational level, WEB-1996 defines regional colleges, which should be autonomous educational institutes organising vocational learning together with local companies. Colleges are responsible for the (quality of) primary learning processes and have to adjust educational programmes to the local needs of companies and the community. Localising the main responsibility for educational quality at college level is a major point of disagreement within the system. The debate since 1996 has focused on quality control mechanisms in VET. The social partners take a prescriptive perspective in this debate: they try to guarantee the quality of vocational courses via qualification structures and examination guidelines, whereas the colleges argue that quality is their responsibility and should be fostered in the interaction between teachers and students. This debate forces a large division in education, with politics determining the social partner position.

6.3. Developing evaluation criteria for VET-systems change

In the Netherlands, the WEB was evaluated politically at the end of 2001, after five years of implementation and use. The Minister for Education established an evaluation committee comprising social partners, representatives of the VET-colleges and scientists and formulated a set of evaluative questions and criteria:

- (a) linking VET and the labour system (the economic demand):
 - (i) is the qualification structure an adequate steering instrument for efficient labour supply (a qualitative instrument to reach quantitative goals!);
 - (ii) the supply and quality of workplaces for learning;
- (b) responding to the social demand: differentiation and quality of educational supply in rela-

- tion to the individual educational demands; establishing basic qualifications; assessing prior knowledge and skills as input instrument;
- (c) quality of educational process and assessment procedures:
 - (i) quality of educational supply (attractiveness of training; consistency between theory-practice-assessment; consistent translation from qualification structure to output);
 - (ii) external verification of assessment and examination procedures;
 - (d) efficiency: internal (time-level ratio; added value) and external (labour-market position) output; efficient internal trajectories;
 - (e) educational system synergy:
 - (i) links towards former education and higher education;
 - (ii) links between adult education and vocational education;
 - (f) autonomy and quality assurance: strategic power of ROCs; knowledge management within ROCs; financial aspects;
 - (g) legal aspects: the role of government; the relation between different actors and stakeholders around VET; supply of public information; (regulation of the) legal position of students.

This set of criteria reflects the history of system development and the actual discussions on education quality: accessibility, efficiency and quality assurance were the red lines for the political evaluation.

The evaluation committee started its work in the summer of 1999. Meetings were held with all stakeholders to enlighten the evaluation task and the (political) quality of the evaluation questions and implicit criteria. To organise a firm, empirical base for the evaluation eight research teams were involved to develop useful sets of measurable variables to assist the political evaluation process (Table 3). The conceptual model 'in use' is a stepping-stone model: the law has impact on institutions and organisations, organisations are in turn responsible for the quality of training and education delivered. In such chain-linked models, lack of quality of the primary process can be caused by autonomous decisions at lower levels and by poor legislation. All political and social stakeholders were involved in the formulation and

implementation of the WEB (the famous Dutch Polder model), so failures in the system are difficult to identify and rectify.

The evaluation model used is highly positivistic: the research teams were intended to take an objective 'picture' of the state of the art in their VET subject. The evaluation was remote from the political and practical debate and was to deliver an objective report to the Minister, who would then make his own political decisions, taking into account or not the results from the evaluation.

6.4. Results and use of the evaluation of WEB-1996

The eight research groups presented their results to the evaluation committee in spring 2001. The evaluation committee then formulated their own conclusions in a separate report presented to education and the Minister for Education during summer 2001. This report was received sceptically, particularly by the social partners (both employers' organisations and trade unions): they did not accept the main conclusion that the quality of VET was the prime responsibility of the regional colleges and not of branch organised bodies and other institutional organisations. In the political battle that followed the publication of the evaluation committee report, scientific and analytical arguments were drawn into the argument. The political conclusion of the Minister was that the evaluation was premature (five years was not long enough to prove labour-market success) and at the same time too late. He decided not to intervene at all, except to introduce new regulations around the examination procedures, establishing a new, regulating institute, prescribing how colleges organise their assessment procedures.

After five years of working under WEB-1996, the results of the evaluation studies are therefore not promising. Although the general quality of Dutch vocational education is high, the results of reaching change goals, as stated in the WEB, are disappointing. The main conclusion in most of the evaluation reports is that the policy concept behind WEB-1996 is no longer compatible with the requirements of a knowledge-based economy. WEB-1996 is built on a strong belief in prescription and forecasting of qualificalational requirements. This belief stems from the industrial work

Table 3: **Evaluation issues, teams, reports and conclusions**

Evaluation theme	Research group	Reports ^(a)	Main conclusions
1A linking VET and labour market	Brandsma (Twente University), de Jong, Karsten and van de Venne	<i>Leren kwalificeren</i> (Learning to becoming skilled)	Linking education and labour market by a qualification structure is not (yet) implemented fully.
1B linking VET and labour market	Heijke (Maastricht University), Borghans, Smit, Hövels and den Boer	<i>De WEB tussen vraag en aanbod</i> (the WEB between demand and supply)	The steering instruments in the WEB are causing a dispersed supply of courses and do not deliver the expected flexibility.
2 responding to social demands	Doets and Westerhuis (Cinop), Klarus, Meijers, Thomas, Vrieze, Dinjens, Neuvel and Pauwels	<i>Voldoen aan de individuele vraag; toegankelijkheid en positie deelnemer</i> (responding to individual demands: openness and individual position)	WEB offers more possibilities for responding to social demands than are used by local actors. The tendency towards autonomy is restricted to lower levels in the system.
3 quality of processes and assessment	Nieuwenhuis (Stoas), van Berkel, Jellema and Mulder	<i>Kwaliteit getoetst in de BVE</i> (the assessment of quality in VET)	Quality assurance is improved in the system, autonomy is restricted at institutional level. Regulating is a Dutch disease.
4 internal and external efficiency	Van der Velden (Maastricht University), Berkenbosch, De Bruijn, de Jong, Voncken, Geerligs and Lokman	<i>Toegankelijkheid, intern rendement en doorstroom</i> (openness, internal efficiency, and destinations)	Lack of reliable data. The system is more accessible, internal efficiency is difficult to estimate, external efficiency and destination is improved.
5 system synergy	De Bruijn (Cinop)	<i>Aansluiting tussen stelsels</i> (linking educational systems)	Cooperation between institutional actors is improved. The organisation of tailor-made trajectories on individual level is still in its infancy.
6 autonomy and quality assurance	Karstanje (Amsterdam University), van Esch, van Ingen, Hoebe, Vermeulen, van de Venne	<i>Zelfsturend stelsel, autonome instellingen, kwaliteitszorg</i> (self steering system, autonomous colleges, quality assurance)	Autonomy and regulations have a tense relationship. WEB offers autonomy, but institutional regulations decrease autonomy. Colleges do not use their options. Deregulation is recommended.
7 legal aspects	Leenknecht (Tilburg University)	<i>Bestuurlijke en juridische aspecten van de WEB</i> (Administrative and legal aspects of WEB)	Division of responsibilities is unclear. Autonomy implies more administration at college level. WEB offers possibilities for realising flexibility at local level.
Overall report	Evaluation committee	<i>De WEB: naar eenvoud en evenwicht</i> (WEB: towards simplicity and equilibrium)	Mixed impressions: WEB has improved VET, but much inflexibility still exists in the system. Institutional and organisational actors are responsible for this inflexibility. Accountability is badly organised. Recommendations are formulated, with the accent on autonomy and professionalism at college level. Institutional organisations should facilitate (not regulate) local educational processes.

^(a) All reports are published (in Dutch) by the Ministry of Education, Culture and Science, in a two-kilo box, together with the final report of the steering committee. Cinop published an English overview edited by Prof. Nijhof of Twente University and Dr van Esch of Cinop (Nijhof and van Esch, 2004).

paradigm, which emerged in coevolution with such Fordist institutional arrangements as full-time employment, clear occupational assignments and a well-established career pattern corresponding to the concept of guaranteed 'job for life'. Young people went to school, got a job and often remained in that job for most of their working lives. Worker networks and trade unions in this system were organised to protect the permanence of these 'life jobs' and to build social welfare programmes around them. Educational systems in this context have been used not only to deliver cognitive skills, but to function as social selectors, steering children from various socioeconomic backgrounds into 'appropriate' levels of education that make them eligible for 'appropriate' jobs. This industrial VET-system could work reasonably well since the hierarchical industrial system was built on jobs that were mostly semi-skilled and changed little over an individual's working life. The system was stratified but could provide security and increasing wages even to those with only basic education (Carnoy and Castells, 1997, p. 36 et seq.; Mayer and Nieuwenhuis, 2001).

In the knowledge-based economy the work paradigm has to change towards the recognition of unstable and unpredictable requirements, which require a different steering concept for VET. The prescriptive educational policies should be altered for a greater reliance on the flexibility and expertise of colleges to organise flexible pathways towards competence in close cooperation with the local companies. The focus of Dutch policies for VET should change from prescription to facilitating accountable VET-professionals and colleges.

At the same time the evaluative studies report a great shortage of expertise within the colleges, both at managerial and teacher levels. The quality of examinations, and therefore the exchange value of qualifications, is dramatically low. Because of the prescriptive nature of WEB-1996, colleges have not been challenged to organise responsiveness and flexibility in their organisations.

In the political debate, both in government and between social partners and other stakeholders,

the main reaction was to establish more detailed prescriptions to improve WEB-1996. The evaluating researchers warn against this counterproductive reaction. The challenge will be to improve WEB-1996 within a forward mode. This will radically change the relationship between institutional and organisational actors; colleges and local companies should be seen as the primary professional actors and the institutional system should be facilitating them instead of setting the rules.

At the beginning of 2002 the Dutch Minister for Education presented his political evaluation of WEB-1996. He largely maintained the status-quo, reasoning that the evaluation studies were either too early (no students had left VET, under the regime of WEB-1996) or too late (regarding the institutions around assessment and examinations). The regulations surrounding examinations have been adjusted according to the critics of social partners, with a more stringent regime and a strong controlling approach.

From an evaluation aspect, the overt use of the results is low. Political arguments in the actual debate have been more decisive than the objective results and opinions of the researchers. The political fight on the executive power on quality regulations was resolved in favour of the prescriptive mode, as stressed by social partners and members of parliament, and the vast number of research reports seem to have been filed away. However, one must not be too pessimistic as the results from the evaluation play a background role in the debate in advisory committees both at political and institutional and organisational levels. The socioeconomic council (in which social partners and government have seats) used the evaluation results in its recent advice on quality and innovation of the VET system (SER, 2002). The results of the evaluation reports are seriously taken into account in the elaborative notes of the Council of national bodies and the Union of regional colleges. A public, learning debate on the perspectives of the Dutch VET system is, however, urgently required.

7. Comparisons and conclusions

7.1. Two approaches to systems change

The two VET systems described are not representative of European diversity in VET. Both systems are essentially school-based, with a dual component built in, comparable to the Nordic and French systems. In the Danish system the dual component of apprenticeship courses is traditionally and culturally more important than in the Dutch system. In the German speaking countries VET is more company based, under the regime of the dual system; whereas the UK systems are more market led. These system differences have an impact on important characteristics such as responsiveness and innovativeness of the systems, as Crouch et al. (1999) have pointed out extensively. Nevertheless, the basic problem with both systems is the same; how to build a flexible VET system, balancing stable structures and dynamic, changing labour-market demands.

In Denmark, a systemic solution is sought for this problem through a widening and flexibilising of educational tracks with broader descriptive frameworks that can accommodate continuous changes in the work system. This would enable students to develop their own occupational identity and competences within the arrangement of the tripartite system: forward mapping and participation in communities of practice is a powerful way of connecting by supporting individual development towards self-employing skills.

In the Netherlands, policy-makers tend to opt for improving industry-based VET. In the Dutch case the institutional connection will be strengthened. From the evaluation reports several significant problems can be foreseen:

- (a) motivational problems for students;
- (b) flexibility problems for local delivery;
- (c) integrative problems for work-based learning in national programmes, the accreditation of prior learning will uncover this rapidly;
- (d) widening discrepancies between exchange value and use value of competences;
- (e) recruitment and motivation problems with VET teachers.

The power of traditional institutions is still quite strong in the Dutch case, connecting school to work in an industrial paradigm through forecasting and prescriptions. However, the Minister for Education in the Netherlands seems to have relaxed the position with (VET) colleges in the last policy document allowing educational content to be chosen by school boards with only a small core curriculum decided upon politically by the government.

Policy-makers are seeking a balance between disconnecting and reconnecting school and work. They are looking for a new paradigm where the locus of connection is moving down from the institutional level towards the primary process of learning and working. At the institutional level this means a shift from a prescriptive policy towards a facilitating policy; at the organisational level this means a shift in organisational culture and professionalism towards autonomy and accountability. In the Danish case, experiments are carried out at a local level. Local economic and educational actors are challenged to organise pathways towards competence development, fitting both the local economic demands as well as the ambitions of future students. Within the Dutch context, experts illustrate the challenges and recommendations for colleges as having to play the role of spiders for innovation networks.

Two different approaches can be seen in these two cases: the Danish approach is characterised by common learning, whereas the Dutch approach is characterised by regulation. Based on Hisschemöller (1993), we can expect the Danish approach to be better than the Dutch. Hisschemöller argues that regulatory policy is only working in situations where agreement exists on values and facts in the problem area. In the industrial society of the former century, the linking of education and labour was seen as a technical problem: facts and values were known, a regulatory solution was suitable. In the merging knowledge-based economy of the 21st century, the predictive power of labour-market forecasting is fading and the debate on employability and life-long learning skills is not finished yet.

Figure 3: Policy instruments

		Values ⇒	
		agreement	disagreement
Facts ⇓	agreement	regulation	pacification
	disagreement	negotiation	learning

Source: Hisschemöller, 1993

According to Hisschemöller, in the case of disagreement on both facts and values, a policy strategy based on common learning is the only effective way to find new solutions, in this case for developing a flexible VET system. All the participants in the Danish reform, including government, have the chance to build their own (hi)stories by trial and error, with features of the new system. In the Dutch case the system is developed on the desk-top, and dropped to the major group of professionals, teachers and in-company masters. In a top-down approach the new system will have difficulty in becoming owned by the shop-floor workers.

The Dutch approach adheres to a directive, linear innovation approach developed elsewhere in concept and design (in industry, this is the research and development department; for education special developmental institutes are involved). In the Danish approach pilot schools, regions, companies and institutional players are involved in the design stage of Reform 2000. All actors are building their own 'me-stories' (what will be my part of the job in the new system) and together are building on 'we-stories' (what it is all about).

The evaluation enterprise is accordingly different in the two approaches: in the Danish case, evaluation is part of the learning process and the evaluators are deeply involved in the evolution of the reform. Evaluation methods, like

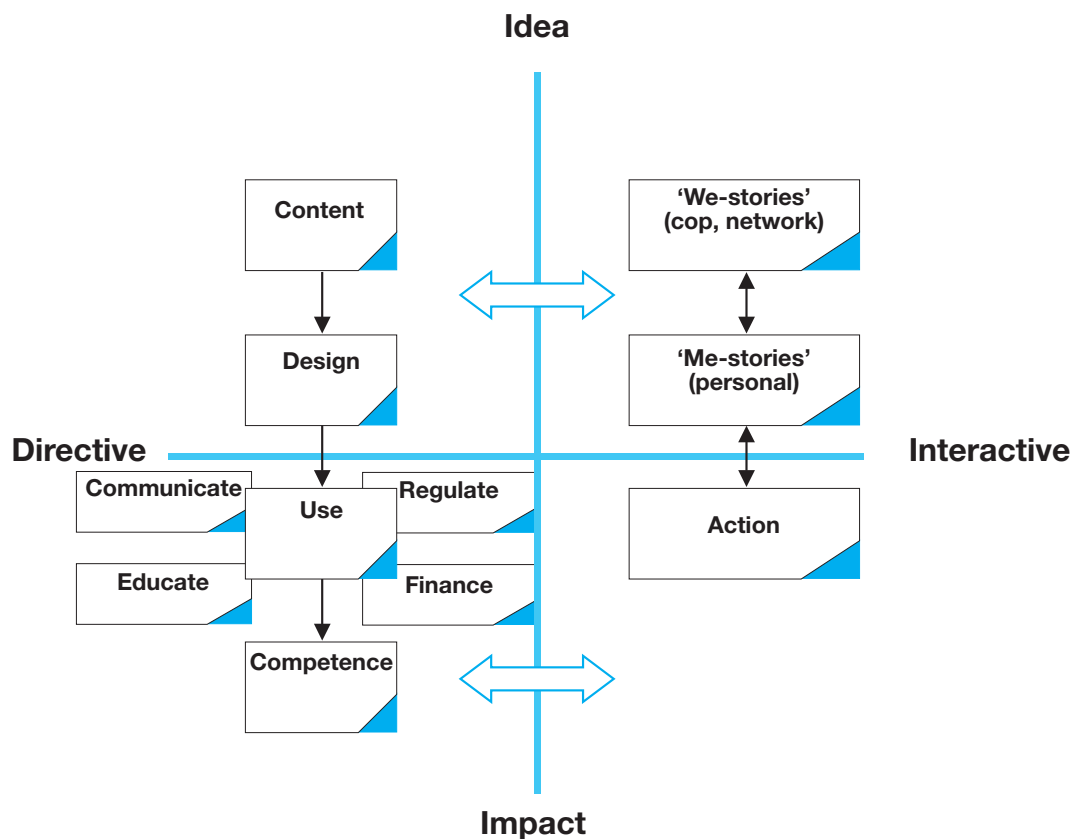
the focus group sessions, are used as implementation instruments to raise awareness of the urgency of Reform 2000. Evaluation is organised alongside the development and implementation of Reform 2000 and the results of the first evaluation round are used immediately for improvement and correction. In the Dutch case evaluation is not used as an accompanying measure for the implementation of WEB-1996, but as an instrument for political control. Evaluation is used more summatively, and the results are (mis-) used immediately in the political debate. So evaluation in the Netherlands did not deliver any common learning amongst the different groups, but made more severe the disagreement and the power game between the ROCs and the social partners.

In the Guba and Lincoln scheme (Section 3) the Danish evaluation can be characterised as participatory, whereas the Dutch evaluation is executed more in the positivistic tradition of remote inquiry and research.

7.2. Implications for pedagogy and didactics in VET

Traditionally, vocational education prepares a student for a particular job based on a body of codified knowledge, an occupational profile with an underlying curriculum basis constructed in

Figure 4: Linear and interactive innovation policies



Source: Geerligs, 2002

terms of detailed objectives and goals. Traditional didactics are reduced to operational lesson plans. The more detailed the planning and safeguards, the better the educational quality. A standard response to new educational demands stemming from societal changes has been to add new themes or even relatively narrowly defined new educational programmes to respond to immediate needs, but not to introduce any major innovation in didactical paradigms and curricular structures. (Abrahamsen and Shapiro, 2001a).

As sectors converge and job profiles and occupations constantly change, spurred on by innovations in technologies and work organisation, the traditions of VET didactics at a policy planning level seem to be becoming obsolete (Grubb, 1999). Nieuwenhuis (2001) suggests a new form of 'participatory didactics', building on the works of Wenger and involving all the institutional actors.

In a recent publication, Christensen (2001) argues that VET didactics is in a major crisis that calls for a revision. On one hand, he advocates that it encompass metadidactics understood as the overall legislative and economic organisation and form that facilitates a pedagogical practice based on learning as construction. On the other hand, he states that this necessitates what he calls 'minimal didactics' to avoid what he sees as 'over-didactisation'. By this he means a didactical practice which both on a policy planning level and at school level is characterised by detailing goals and objectives, leaving no space in the design for active learning participation and joint creation.

The pedagogical foundation for the VET reform (Shapiro and Christensen, 1999) proposes a change towards a broader didactic framework that is constituted in a constructivist view of learning. It is built on a three layer basis that encompasses institutional actors at different

levels, trying to take a minimalist approach in developing a descriptive framework supported by a consistency in terminology at all levels. The aim has been to provide a consistency between the narrative of a new didactic space and the legislative basis that introduces and legitimises some instruments for change.

7.3. Towards general European conclusions

Most European VET systems are in a period of transition searching for new equilibria in an economy where learning plays a growing role in economic performance and societal sustainability. In this context, VET policies play a central role in shaping efficient innovation systems that can adapt to rapid change.

In research on the creation of adaptable innovation systems three inter-related issues have been identified:

- (a) stimulating learning institutions and economic actors;
- (b) developing integrative policy visions and instruments for enhancing innovation;
- (c) creating conditions for policy-making processes which are constantly learning and adapting to new demands and conditions (Lundvall and Borrás, 1997).

Studies on the nature of policy change have traditionally taken their point of departure in the policy-cycle where the policy process is analysed as set in different distinct stages: decision-making, implementation and evaluation. The contracted study on the evaluation of Danish and the Dutch VET reform reflects this view to a certain degree. The learning approach as brought forward by researchers such as Lundvall and Borrás criticises this assumption because it does not seem to provide a thorough account of what happens after the decision-making phase and it tends to perceive change as something automatic following the political impact of the evaluation (Lundvall and Borrás, 1997).

The learning approach provides a more fluid perspective on the policy process in continuous transformation and evolution where no clear stages can be discerned.

'In the political environment of public management learning, processes are particularly difficult

to create and maintain. Individual learning is a psychological process. Organisational learning is a political process. A critical task of public management is to build institutional learning capabilities at the macro level to manage the environment in which private management operates. Conventional policy processes often block learning because ideology overrides evidence or vested interests resist policy evaluation and change ... public management is better regarded as management by design than by direction. Therefore, policy-makers should be concerned with designing adaptable systems rather than producing blue-prints for specific reforms.' (Metcalfe, 1993; quoted by Lundvall and Borrás, 1997).

In the learning approach the key to understanding policy change and the implicit learning dimension of the policy process are transformations in the collective beliefs within the broader system of actors (Sabatier and Jenkins-Smith, 1993). From this perspective the evaluation is no longer just a matter of tracking progress against the desired policy goals. One of the real challenges, which was quite evident from the first and the second evaluation in Denmark, is for the evaluation to be conducted and communicated in such a form that it creates a 'stage' that can invoke a new narrative to replace old terminology and through that replace old practices (i.e. the classroom, the weak achiever) to guide innovation in a shared and intended direction (Shapiro et al., 2001).

The contexts in which both the Danish and the Dutch reform were originally set and implemented, demonstrate an emerging need to develop methodologies which can proactively address the formulation of VET policy in more fluid and dynamic ways and in ways that are more integrated with other policy priorities. Fundamentally such approaches would break with the policy planning and implementation approaches of the past. Policy formulation is understood as a contingent policy development strategy. Incremental innovations should be allowed through the creation of continuous feedback mechanisms from the field of activity in ways that can be processed to enable real-time knowledge-based modifications of policy. When such approaches are achieved, they would be recognised as being appropriate for contempo-

rary times of turbulent change (Gavigan, 2000). In contrast, the more static approach to VET evaluations would typically be dependent on multi-annual periods of relative stability. This seems to have been the underlying understanding behind earlier reforms and subsequent evaluation strategies.

As we have argued, changing VET-systems can be seen as a kind of governmental learning. Perceptions of policy process and subsequent formulation of policy strategies and targets of systems change should be defined based on the specific problem definition and the specific configuration of institutional and organisational actors and their stakes. The complexity of actors involved combined with the speed, uncertainty and the complexity of roles that VET is supposed to fulfil in a knowledge-based economy would favour such a learning approach for governance. Reform of European VET policies should not be sought in prescriptive regulations. The most successful road for European policies on VET is a set of challenging goals and evaluation criteria whose targets should be dealt with in the frame-

work of the dilemmas described in this article. Also at systems level a backward mapping, blueprinting policy is inferior to a forward-mapping developing concept of systems evolution. Challenging evaluation criteria can form the core of a development-oriented European VET policy. These criteria should be derived from good practices: benchmarking is the way European policies for VET should develop.

The emergent knowledge-based economy requires local and temporary solutions: national and European policies should deliver the preconditions for these solutions but should try to avoid prescribing solutions. Similarly with national bodies of social partners: they can challenge colleges and companies to deliver local solutions for global problems and they can facilitate this process but they cannot efficiently prescribe. The one-and-only right solution does not exist and some problems are even unsolvable. This implies a learning approach for politics embedded in broader and integrated innovation policies, an approach, which is yet in its infancy (Van der Knaap, 1997; Gavigan, 2000).

List of abbreviations

ICT	information and communication technology
ROC	<i>Regionaal opleidingencentrum</i> [regional training centres]

References

- Abrahamsen, M.; Shapiro, H. *IT udviklingen og nye lærer kvalifikationer og curriculum i Ingeniøruddannelserne*. Aarhus: Danish Technological Institut, 2001a (Analyse for IDA).
- Abrahamsen, M.; Shapiro, H. *På sporet af nye lærer kvalifikationer i lyset af IT udviklingen*, Aarhus: Danish Technological Institut, Kompetence and IT, 2001b.
- Argyris, C.; Schön, D. *Organisational learning: a theory of action perspective*. Reading, Mass: Addison Wesley, 1978.
- Argyris, C. *On organizational learning*. Cambridge: Blackwell Publishers, 1992.
- Bale, L. S. *Gregory Bateson's theory of mind: practical applications to pedagogy*, 1988. Available from Internet: http://www.narberthpa.com/Bale/lsbale_dop/learn.htm [cited 11.9.2003].
- Bloom, B. *Taxonomy of educational objectives*. New York: Longman, 1956.
- Boaz, A.; Hayden, C. Proactive evaluation: enabling research to be useful, usable and used. *Evaluation*, 2000, Vol. 8, No 4.
- Boaz, A. *National evaluations – local evaluators*. UK Centre for Evidence-based Policy and Practice. UK Evaluation Society Conference December, 2002.
- Brown, J. S.; Duguid, P. Organizational learning and communities of practices: towards a unified view of working, learning and innovation. *Organizational Science*, 1991, Vol. 2, Issue 1.
- Caplan, N. The two communities; theory and knowledge utilisation. *American Behavioral Scientist*, 1979, Vol. 22, Issue 3, p. 459-471.
- Carnoy, M.; Castells, M. *Sustainable flexibility*. Paris: OECD, 1997.
- CERI. *Educational research and development: trends, issues and challenges*. Paris: OECD, 1995.
- Christensen, F. *The important and innovative features of the Reform 2000 of the Danish VET system*. Copenhagen: Ministry of Education, 1999 (working paper).
- Christensen, F. *Didaktikkens tre aktuelle kriser. Uddannelse, læring og Demokratisering*. Copenhagen: Danish Ministry of Education, 2001.
- Christensen, F.; Shapiro, H.; Kjær, F. *Pædagogiske og didaktiske overvejelser bag erhvervsuddannelsesreform 2000*. Copenhagen: Ministry of Education, 2000 (Uddannelsesstyrelsens temahæfteserie, No 34).
- Coffield, F. 101 initiatives but no strategy: policy on lifelong learning in England. In: Nieuwenhuis, L.; Nijhof, W. J. (eds) *The dynamics of Vet and HRD systems*. Enschede: Twente University Press, 2001.
- Collins, A; Brown, J. S.; Newman, S. E. Cognitive apprenticeship: teaching the craft of reading, writing and mathematics. In: Resnick, L. B. (ed.) *Knowing learning and instruction*. Hillsdale, NJ: Erlbaum, 1989.
- Collins, P. H. *Black feminist thought. Knowledge consciousness and the politics of empowerment*. New York: Routledge, 1990.
- Crouch, C.; Finegold, D. Sako, M. *Are skills the answer; the political economy of skill creation in advanced industrial countries*. Oxford: Oxford University Press, 1999.
- Czarniawska, B. *The uses of narrative in social science research*. Göteborg: School of Economics and Commercial Law, 2001.
- Danish Ministry of Education. *Regelstyring i Erhvervsuddannelserne, Rapport af 19. maj 1998*. Copenhagen: Danish Ministry of Education, 1998.
- De Bruijn, E. *Het experimentele en het reguliere; twintig jaar voltijds KMBO. [20 years of school based VET]*. Amsterdam: PhD study University, 1997.
- Denzin, K. The practices and politics of interpretation. In: Denzin, N.; Lincoln, Y. S. (ed.) *Handbook of qualitative research. Second edition*. Thousand Oaks: Sage Publication, 2000.
- 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.
- Ellström, P. E. *Integrating learning and work: problems and prospects*. Linköping: University of Linköping, 1999 (Forum Paper).
- European Commission, DG XXII. *Teaching and learning: towards the learning society: white*

- paper on education and training*. Luxembourg: Office for Official Publications of the European Communities, 1995.
- Finegold, D. Institutional incentives and skill equilibrium. In: Ryan, P. (ed.) *International comparisons of vocational education and training for intermediate skills*. London: Falmer Press, 1991.
- Foray, D.; Grübler, A. Technology and the environment: an overview. *Technological Forecasting and Social Change*, 1997, Vol. 53, No 1, p. 3-13.
- Gardner, H. *De mange intelligensers pædagogik*. Copenhagen: Gyldendal Undervisning, 1997.
- Gavigan, J. *Knowledge and learning, the futures project*. Sevilla: Institute for Prospective Studies, 2000.
- Geerligs, J. *Innoveren in de greep. [Catching innovation]*. Wageningen: Stoas, 2002.
- Grubb, W. N. *Conceptualising vocational education*. Berkeley: Institute of Education, 1999 (working paper).
- Guba, E. G.; Lincoln, Y. S. *Fourth generation evaluation*. Newbury Park: Sage Publications, 1989.
- Hanusch, H.; Shionoya, Y. *Innovations in technology and institutions. Studies in Schumpeterian perspectives*. Ann Arbor: University of Michigan Press, 1994.
- Hisschemöller, M. *De democratie van problemen. [The democracy of problems]*. Amsterdam: VU-press, 1993 (PhD study).
- IPTS. *The new economy*. Spain: IPTS, 2000 (working paper).
- Jonassen, D. Evaluating constructivist learning. *Educational Technology*, 1991a, Vol. 36, Issue 9.
- Jonassen, D. Objectivism vs constructivism. *Educational Technology and Research*, 1991b Vol. 39, Issue 3.
- Jonassen, D. Thinking technology. *Educational Technology*, 1994, Vol. 34, Issue 4.
- Kelleher, M. *VET as a learning organisation*. Usk: CRED, 2001 (Working paper, Forum Network).
- Kelleher, M.; Hughes, J. Evaluating VET from a learning organisation perspective. (Paper for the research forum on VET). Wales: CRED, 2000.
- Kvale, S.; Nielsen, K. *Læring i moderne dansk vekseluddannelser Forprojekt*. Aarhus: Psykologisk Institut, 1999.
- Lave, J.; Wenger, E. *Situated learning, legitimate peripheral participation*. Cambridge: University press, 1991.
- Levin-Rozalis, M. Abduction: a logical criterion for programme and project evaluation. *Evaluation*, 2000, Vol. 6, No 4, p. 415-532.
- Lincoln, Y. S.; Guba, E. G. Ethics and politics in qualitative research. In: Denzin, N.; Lincoln, Y. S. (ed.) *Handbook of qualitative research. Second edition*. Thousand Oaks: Sage Publication, 2000, p. 163-188.
- 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*. Luxembourg: European Commission, TSER, 1997.
- Mayer, K.; Nieuwenhuis, L. The competence of VET for the knowledge based economy. In: Achtenhagen, F.; John, E. G. (eds) *Milestones of vocational education and training: institutional perspectives*. Bielefeld: W. Bertelsmann Verlag, 2001.
- Metcalfe, S. J. Public management: from imitation to innovation. In: Kooiman, J. (ed.) *Modern governance: new government society interactions*. London: Sage, 1993.
- Neisser, U. *Cognition and reality*. San Francisco: Freeman & Co., 1976.
- Nielsen, K.; Kvale, S. (eds) *Mesterlære – læring som social praksis*. Copenhagen: Hans Reitzel, 1999.
- Nieuwenhuis, L. *Change in VET: a systems' approach*. Wageningen: Stoas Research, 2001 (Working paper Forum Network).
- Nieuwenhuis, L. F. M. et al. *Kwealiteit getoetst in de BVE [Testing quality in VET]*. Zoetermeer: Ministry of Education and Science, 2001.
- Nijhof, W. J. Transformaties in het beroepsonderwijs [Transforming VET]. In: Nijhof, W. J. (ed.) *Ontwikkelingen in het beroepsonderwijs en de volwasseneneducatie. Onderwijskundig lexicon*. Alphen aan de Rijn: Samson H. D. Tjeenk Willink, 1997.
- Nijhof, W. J. The VET system between private demands and public interests, the Dutch challenge. In: Nieuwenhuis, L.; Nijhof, W. J. (eds) *The dynamics of Vet and HRD systems*. Enschede: Twente University Press, 2001.
- Nijhof, W. J.; van Esch, W. (eds) *Unravelling policy, power, process and performance. The forma-*

- tive evaluation of the Dutch adult and vocational education law*. Den Bosch: Cinop, 2004.
- Nonaka, I. *The knowledge creating company*. Oxford: University Press, 1995.
- OECD. Ministers for Education communiqué on high quality education and training. In: CERI. *Educational research and development: trends issues and challenges*. OECD: Paris, 1995.
- OECD. *Pathways and participation in vocational and technical education and training*. Paris: OECD, 1998.
- Orr, J. Sharing knowledge, celebrating identity, war stories and community memory in a service sector. In: Middleton, D. S.; Edwards, D. *Collective remembering, memory in society*. Beverly Hills, Call: Sage, 1990.
- Resnick, L. B.; Wirt, J. G. (eds) *Linking school and work: roles for standards and assessment*. San Francisco, CA: Jossey-Bass Inc., 1996 (ED 389 915).
- Rossi, P. H.; Freemann, H.; Lipsey, M. W. *Evaluation a systemic approach*. Thousand Oaks. Sage, 1999.
- Sabatier, P. A. (ed.) *Theories of the policy process*. Boulder, CO: Westview Press, 1999.
- Sabatier, P. A.; Jenkins-Smith, H. C. (eds) *Policy change and learning: an advocacy coalition approach*. Boulder: Westview Press, 1993.
- Sanderson, I. Evaluation in complex policy systems. *Evaluation*, 2000, Vol. 6, No 4, p. 433-454.
- Schön, D. The reflective practitioner – how professionals think in action. New York: Basic Books, 1983.
- SER – Socio-economic Council. *Koersen op vernieuwing. [On innovation]*. The Hague: SER, 2002.
- Shapiro, H. *Flatland revisited – voyage through a changing learning landscape*. Presentation at International Community College Conference in New Orleans, organised by Regional Technology Transfer and USA Department of Education, 2002.
- Shapiro, H. *Knowledge show me the way*. North Carolina, 2000 (Conference proceedings).
- Shapiro, H.; Abrahamsen, M. *Erfaringer fra Forsøg Med EUD Reformen. Første og Snden Erfaring-sopsamling*. Copenhagen: Danish Ministry of Education, 2001 (Uddannelsesstyrelsens temahæfter).
- Shapiro, H.; Christensen, F. *Pædagogisk Grundlagsnotat, Arbejdsnotat*. Copenhagen: Danish Ministry of Education, 1999.
- Shapiro, H. et al. *Erfaringer med Erhvervsuddannelsesrefomen- anden erfaringsopsamling*, Copenhagen: Teknologisk Institut, 2001 (Uddannelsesstyrelsens temahæfter).
- Shapiro, H. et al. *Forsøg med erhvervsuddannelsesreformen- Opsamling af de Første erfaringer efteråret 1999*. Copenhagen: Danish Ministry of Education, 2000 (Uddannelsesstyrelsens temahæfter).
- Stasz, C. Assessing skills for work: two perspectives. *Oxford Economic Papers*, 2001, Vol. 53, Issue 3, p. 385-405.
- Stern, D. Institutions and incentives for developing work-related skill. In: Adler, P. (ed.) *Technology and the future of work*. Oxford: Oxford University Press, 1992, p. 49-186.
- Stuurgroep Evaluatie Web. *De Web: naar eenvoud en evenwicht [WEB: towards simplicity and equilibrium]*. Zoetermeer: Ministry of Education and Science, 2001.
- Sutton, R. *The policy process: an overview*. London: Overseas Development Institute, 1999.
- Ulriksen, L. *Almenkvalificeringsprojektet – Perspektiver på almenkvalificering*. Roskilde: Roskilde Universitet, EVU Gruppen, 1992.
- Van der Knaap, P. *Lerende overheid, intelligent beleid [Learning government; intelligent governance]*. Amsterdam: Free University, 1997 (PhD thesis).
- Van der Meer, F. B. Evaluations and the social construction impacts. *Evaluation*, 1999, Vol. 5, No 4, p. 387-406.
- Weiss, C. The many meanings of research utilization. *Public Administration Review*, 1979, Vol. 39, Issue 5, p. 426-431.
- Wenger, E. *Communities of practice, learning, meaning and identity*. New York: Cambridge University Press, 1998.
- Zuboff, S. *In the age of the smart machine: the future of work and power*. New York: Basic Books, 1988.



Evaluation of EU and international programmes and initiatives promoting mobility: selected case studies

Wolfgang Hellwig, Uwe Lauterbach,
Hermann-Günter Hesse, Sabine Fabriz

Abstract

International exchange and promotion programmes began to become relevant to vocational education and training after World War II, initially operating at transnational level between nation states. In Europe the binational or multinational programmes of nation states or specific providers have been gradually complemented or replaced by those of supranational institutions, such as EU programmes. Programmes have also been developed by national and international organisations (such as the World Bank, the OECD and the ILO) and by supranational institutions (e.g. the EU) for vocational training cooperation with threshold and developing countries.

Those arranging and funding these programmes, particularly at a political level, want programme evaluation to provide feedback and explain and justify public expenditure. They also expect continuous improvement of programmes through more precisely defined goals.

Formative evaluation requires an approach capable of explaining the origin of desired or undesired results more so than summative evaluation. At a teaching level, the explanation is usually complex. Only a sufficiently complex explanatory model can provide valid suggestions for improvement. Consequently, evaluation must build on a design that takes potential success factors into account. However, an evaluation design must not only fulfil the criteria of complexity and precision, but it must also be acceptable to trainees and project organisers and economically viable. Because of the high costs involved, it is rare for an evaluation to be carried out to a high scientific standard. It is important, however, to reach an acceptable compromise. To help practitioners decide on the appropriate approach and on what are relevant results, a description of evaluation tools and their relative advantages and disadvantages are included here.

After an introduction focused on evaluation standards we selected four examples: Leonardo da Vinci Programme of the EC, the Community Study Visits Programme of the European Commission, the German-French Youth Foundation and the International Exchange Programme for VET Specialists of the German Federal Government. A summary and recommendations for future practice conclude this study.

Table of contents

1. Premises and objectives of the study	97
1.1. Scope of study and evaluation	97
1.2. Reasons for the selection of case studies and their subdivision	97
1.2.1. Promotion of mobility	97
1.2.2. Programme goals and providers	98
1.2.3. Programme participants	98
1.2.4. Programme structures and programme organisation	98
1.2.5. Programme providers	98
1.2.6. Selection of case studies and evaluation research	99
1.2.7. Cases chosen	99
1.2.8. Structure of case studies	100
1.3. Evaluation standards and appraisal of EU, transnational and international programmes	100
1.3.1. Significance of evaluation standards for the appraisal of EU programmes	100
1.3.2. Criteria for programme evaluation	101
2. Case study 1: the LdV programme (mobility)	106
2.1. General framework	106
2.2. Programme objectives	107
2.3. Evaluation of the overall programme, individual programmes, particular aspects, etc.	108
2.3.1. Overall programme	108
2.3.1.1. Implementation of the programme in the EU	108
2.3.1.2. Implementation of the programme in Germany	110
2.3.2. Evaluation of individual measures and particular aspects, in Germany and elsewhere	113
2.3.2.1. LdV evaluation practice at the ZAV	113
2.3.2.2. LdV evaluation practice at InWent	115
2.4. Summary	115
3. Case study 2: European Community Study Visits programme for those responsible for Vocational Training (CSV)	118
3.1. General framework	118
3.2. Programme objectives	118
3.2.1. Programme participants	119
3.2.2. Programme procedure and content	119
3.2.3. Peculiarities, reporting	120
3.3. Evaluation of the programme as a whole, individual programmes, particular aspects, etc.	120
3.3.1. Programme dimension, evaluation stakes and documentation of findings	120
3.3.2. Scope of the evaluation and focal points	121
3.3.3. Evaluation material, data, procedures and costs	122
3.3.4. Examples of evaluation	122
3.3.4.1. National evaluation (Germany 1995-98)	122
3.3.4.2. Overall evaluation of CSV	127
3.3.4.3. Measures to support (continuing) evaluation	128
3.4. Summary	128
4. Case study 3: the German-French Youth Foundation (DFJW)	130
4.1. General framework	130

4.2. Programme objectives	130
4.3. Evaluation of the overall programme, individual measures, particular aspects, etc.	131
4.3.1. Internal evaluation analysis	132
4.3.1.1. Intercultural relations and intercultural learning study conference	132
4.3.1.2. Acquiring skills for Europe	133
4.3.2. Future planning	134
4.4. Summary	134
5. Case study 4: International Exchange Programme for VET Specialists (IFKA)	136
5.1. General framework	136
5.2. Programme goals	137
5.2.1. Objectives	137
5.2.2. Participants	137
5.2.3. Use of electronic media for preparation, implementation, dissemination and lasting effect	137
5.2.4. Partner countries	137
5.3. Evaluation of the programme as a whole, individual programmes, particular aspects, etc.	138
5.3.1. Principles and focuses	138
5.3.2. Chronology of the German Institute for International Educational Research (DIPF) evaluation of IFKA	138
5.3.3. Interdisciplinary team of evaluators	146
5.3.4. Dissemination of findings	146
5.4. Summary	146
6. Summary and recommendations for future practice	148
6.1. Four practical models	148
6.2. Implementation of evaluation and practical examples	148
6.2.1. Programme organisation, evaluation planning and preparation	149
6.2.2. Programme goals, development of suitable methods and tools for evaluation	150
6.2.3. Quality of evaluation, optimisation potential	150
6.2.4. Transformation of evaluation results and financial framework	152
6.3. Prospects for a new evaluation culture and 12 proposals for the implementation of evaluation	153
List of abbreviations	155
References and bibliography	157

List of tables and figures

Tables

Table 1:	Applications for LdV mobility projects: target groups and implementation authority	113
Table 2:	Raster for assessment of the most important project phases of LdV mobility projects from the point of the participants	114
Table 3:	The CSV: figures for Germany 2000-02 (incoming, outgoing)	119
Table 4:	Evaluation of the CSV referred to Germany 1995-98: research tools and data sets	123
Table 5:	Checklist for assessment of participants' reports	124
Table 6:	1979-83: documentation of findings and formative evaluation	138
Table 7:	1983-94: formative evaluation and documentation of findings	139
Table 8:	1983-96: summative evaluation, added value of IFKA, formative proposals for programme redesign	140
Table 9:	1995-2001: annual formative evaluation	145
Table 10:	Mobility in VET – one of many ways of moving towards the principal objective	148
Table 11:	Mobility in VET – two concrete measures	148

Figures

Figure 1:	Evaluation structures of the LdV mobility programme	117
Figure 2:	2001 CSV evaluation structure before introduction of new evaluation practices in 2002-03	128
Figure 3:	DFJW evaluation structures	134
Figure 4:	Structural analysis of evaluation (survey of participants) 1984-96	142
Figure 5:	Schematic model of 'programme success' factors, taken from surveys of participants and assessment of participants' reports	144
Figure 6:	IFKA evaluation structures	147

1. Premises and objectives of the study

1.1. Scope of study and evaluation

The tradition of international exchange and promotion programmes goes back more than a century. Their particular relevance to vocational education and training (VET) began after World War II. Initially they functioned at transnational level between nation-states.

The binational or multinational programmes of nation-states or specific providers (such as foundations, national and multinational societal institutions, etc.) have gradually been complemented or replaced by those of supranational institutions, such as European Union (EU) programmes within Europe.

The programmes mentioned above relate to industrialised countries (including those outside Europe) and the geopolitical area of Europe in its broadest sense. Programmes have also been developed by national and international organisations (such as the World Bank, the OECD and the ILO) and by supranational institutions (e.g. the EU) for vocational training cooperation with threshold and developing countries.

Those arranging and funding the programmes, particularly at a political level, want programme evaluation to provide feedback and explain and justify public expenditure. They also expect continuous improvement of programmes through more precisely defined goals.

Formative evaluation requires an approach capable of explaining the origin of desired or undesired results even more so than summative evaluation. At a teaching level, the explanation is usually complex, that is, we may presume that groups of factors are responsible for a given result and may replace or compensate for one another. Only a sufficiently complex explanatory model can provide valid suggestions for improvement. For this reason, evaluation must build on a design that takes potential success factors into account. However, an evaluation design must not only fulfil the criteria of complexity and precision, but must also be acceptable to trainees and project organisers and economically viable. It is

rare for an evaluation to be carried out to a high scientific standard (control group design, reliable and construct-valid data capture, consideration of many factors), because of the high costs involved. What is more important is reaching an acceptable compromise. To make practitioners aware of this and enable them to make an informed decision on the approach to take and what results are relevant to their individual context, we consider it worthwhile to offer a description of evaluation tools, including their relative advantages and disadvantages.

1.2. Reasons for the selection of case studies and their subdivision

1.2.1. Promotion of mobility

Examples of evaluation practices for mobility programmes are examined at the following levels:

- (a) at an EU level, which programme organisers expect to be directly beneficial to fostering the European dimension of VET. Not only supranational programmes promote international skills, such as mobility and intercultural competence or the international political dimension.
- (b) At a national level, mobility programmes with a binational or international dimension have also been developed. Their programme structures, which are less complicated than those of the EU, render them entirely appropriate as case studies. Furthermore, because of continuity and long experience, the considerable expertise of those involved in the evaluation of programmes and projects (measures) can be exploited.

The four case studies from the above geopolitical areas, contrasted with an 'ideal evaluation', will constitute the basis for a scientific conclusion to this study. We shall focus on applicability, in order to offer EU-programme organisers pointers for their own programmes and projects.

1.2.2. Programme goals and providers

In order to identify potential case studies more accurately in the complex field of the EU and international, binational and national VET providers, we need to distinguish between measures which:

- (a) aim directly at promoting development or (re)structuring of VET in a State or region, etc. Typical programmes are those of the European Social Fund (ESF), development cooperation and World Bank projects in transitional countries;
- (b) intend to generate an indirect impact on the development of VET by creating 'special scenarios' to be used by programme participants, as in the case of the Leonardo da Vinci (LdV) ⁽¹⁾ programmes to promote vocational mobility in the EU, such as the Community study visits programme for those responsible for vocational training (CSV) and national programmes to promote international training of VET staff, such as the international exchange programme for VET specialists (IFKA) in Germany;
- (c) set out to generate an indirect impact on binational relations and promote 'eurocapability' and international skills, such as the German-French Youth Foundation, *Deutsch-Französisches-Jugendwerk*, DFJW) and support these programme goals with a number of measures from all spheres of life. These give particular importance to vocational training for the target group.

In addition to the character-forming of programme participants from VET, the transfer of 'best practice' or 'good practice' and the lending, borrowing and adaptation of these to particular scenarios play a significant role in the indirect (b and c) and direct (a) promotion of VET (Lauterbach, 2003, p. 108 et seq.).

This brief excursion into possible programme aims and structures, dissemination, sustainability and evaluation of results already shows that there

are any number of special cases within EU and international programmes.

1.2.3. Programme participants

Core VET participants comprise:

- (a) partners and social partners in enterprises, business and educational establishments;
- (b) apprentices/trainees/students/interns and teaching personnel (teachers, trainers, etc.);
- (c) the framework designers and administrators such as policy-makers, educational administrators, providers bodies and other institutions;
- (d) researchers and evaluators from academia, research and consulting.

1.2.4. Programme structures and programme organisation

Programme organisation creates a range of programme structures. These are characterised by such features as:

- (a) programme duration (short, medium or long term);
- (b) individual and group programmes;
- (c) homogeneity, or the lack of it, in certain characteristics;
- (d) reciprocity, or the lack of it, in programmes;
- (e) direct incorporation into host country institutions (e.g. company or school placements, projects, higher education) or 'external' observation, analysis, visits and counselling.

1.2.5. Programme providers

Case studies included here are not only determined by the above criteria mentioned in Sections 1.2.1 to 1.2.4 but also by access to the programme providers, the programme organisers. The remarkable significance of the LdV programme requires the identification of the organisers of its projects. Chief among these in Germany is InWent ⁽²⁾ which, for more than 20 years, has been the principal provider of initial training schemes abroad. The LdV experience of

⁽¹⁾ Since CSV falls within LdV, reference should be made to the EU Community action programme on vocational training, LdV phase II, 2000-06. In this the European Commission lays down such matters as the content of mobility schemes: 'Promotion of cross-border mobility for people in vocational training, particularly young people and VET decision-makers' (Mobility); L146/40, 11 June 1999 (More information on LdV in Chapter 2).

⁽²⁾ InWent – *Internationale Weiterbildung und Entwicklung gemeinnützige GmbH* (Capacity Building International, Germany) is an organisation for international human resources development, advanced training and dialogue. It was established in November 2002 through a merger of *Carl Duisberg Gesellschaft e.V.* and the *Deutsche Stiftung für Internationale Entwicklung* (German Foundation for International Development).

the German academic exchange service (*Deutscher Akademischer Austauschdienst*, DAAD) and the Central placement office (*Zentralstelle für Arbeitsvermittlung*, ZAV which is responsible for placement of young employees should also be incorporated in the study. CSV, organised centrally by the European Centre for the Development of Vocational Training (Cedefop), is also administered in Germany by InWent. A wide range of programme evaluation literature on DFJW schemes is available.

1.2.6. Selection of case studies and evaluation research

Selection of case studies, based on modern evaluation research, should pay particular attention to certain key topics. Goal definition and coherence of project design must be verified. We should particularly guard against aiming to meet too many goals at too high a level in too short a time. Secondly, the goals chosen should be subdivided into interim objectives to be attained in ascending order of difficulty, reflecting participants' initial circumstances.

1.2.7. Cases chosen

The four cases selected were:

- Case study 1: LdV action programme: mobility;
- Case study 2: CSV – Community study visits programme (henceforth part LdV);
- Case study 3: the German-French Youth Foundation (*Deutsch-Französisches Jugendwerk*, DFJW);
- Case study 4: the international exchange programme for VET specialists (IFKA).

They represent different types of programme, each corresponding to a greater or lesser extent to the criteria given in Sections 1.2.1 to 1.2.6. All fulfil criterion 'mobility' (Section 1.2.1). The promotion of vocational training mobility is the key programme goal. All possible aspects of individual promotion and 'system mobility' in VET are incorporated in case study 1. In case studies 2 and 4 individual involvement is foremost, but the promotion of system and project mobility is also a necessary consequence. Case study 3 relates the issue of individual mobility to the political goals of

promoting international understanding. Ultimately, these underlie all the case studies.

Since LdV provides the framework for many measures ⁽³⁾, almost the entire range of criteria such as 'programme participants', 'programme structures and organisations' and 'programme providers' are covered. With regard to the criterion 'case studies and evaluation research', we shall need to verify whether the 'mobility' factor can be independently evaluated within LdV, or whether some other path might be followed to carry out evaluations within this complex programme structure.

Case studies 2, 3 and 4 are characterised by clearly defined goals. Programme goals correspond to criterion 'case studies and evaluation research'. Like case study 1, they relate to indirect effects in the promotion of VET mobility (Section 1.2.2, b and c). They differ in programme providers and the composition of the individual measures (mono-national, binational or multinational) in programme structures and programme organisation.

Access for our research team programme providers and the various levels of evaluation hitherto completed, were also key factors in the choice of case studies, which differ in their target groups but all concern VET and mobility.

A complete evaluation of case study 1 is virtually impossible. (In this case evaluation is limited by the budget available, notwithstanding expediency concerns over its usefulness.)

Case study 2 offers a plethora of evaluation results but these differ greatly in quality. The development of a culture of continuing evaluation has received particular attention since 2000 as the framework has expanded.

Case study 3 demonstrates broad experience of evaluation, particularly in the general context of binational exchanges. This basis retains its validity in the particular case of VET.

Case study 3 is of particular interest to our project for another reason. The entire DFJW does not attain the dimensions of LdV (case 1) but, as a meta-programme, its structure is similar. Case study 3 shows how evaluations of framework programmes containing many subprogrammes, individual measures, particular aspects, etc., may be conducted to good effect.

⁽³⁾ See Chapter 2 for programme description.

Just as cases 1 and 3 resemble one another, so do 2 and 4 ⁽⁴⁾. The aims and structures of the individual programmes, containing many annual activities, are similar. The above-mentioned difference between mono and multinational levels does not in fact influence evaluation design. Case study 4 was included because it represents more than two decades of evaluation culture, which enables it to apply complex social science processes, such as path analysis, as tools of evaluation. This continuity has also led to the incorporation of programme evaluation in programme planning and operations. The dissemination of evaluation results (publications, events) related to programme findings was another factor in the choice of case study 4.

1.2.8. Structure of case studies

The same basic form of reporting was employed for each case study:

- (a) programme objectives;
- (b) evaluation of the programme as a whole, of individual measures, particular aspects, etc.;
- (c) summary;
- (d) literature, sources, etc.

Subsections to clarify certain aspects are included where necessary as are descriptions of the evaluation measures in Chapters 2 to 4.

1.3. Evaluation standards and appraisal of EU, transnational and international programmes

1.3.1. Significance of evaluation standards for the appraisal of EU programmes

The more an evaluation is intended to generate knowledge that can be used for intervention to improve a programme and take stock of it, the

greater the need for it to be guided by and linked to scientific research methods. In contrast to surveys, if change is to be scientifically based, it is not sufficient for it to deal with obvious findings. The knowledge required to make changes requires research on implementation, intervention and dissemination. This must systematically investigate how programmes fare in practice and the extent to which theoretically expected effects may be empirically confirmed under quasi-experimental controlled conditions.

In the light of recent trends in evaluation research and with particular regard to the debate on the international educational achievement assessments TIMSS, PISA, IALS and the new DESI ⁽⁵⁾, we take evaluation to be a scientific endeavour in which premises, processes and effects of programmes are considered and compared with stated goals and evaluation and research are explicitly linked. Evaluation is as much subject to the methodological standards of scientific activity as is research.

Evaluation models should take the theoretical and methodological state of knowledge of relevant disciplines as a starting point and make a general contribution to research beyond the immediate scope of their own concerns, so as to participate in the general future development of evaluation standards. Unlike research, evaluation projects serve as bases for institutional decisions. They are oriented to the questions, goals and appraisal criteria of stakeholders and should be of use to them. Therefore, the stipulations we adopted were:

- (a) evaluations should be scientifically based empirical undertakings that set out hypotheses and appraise the design, implementation, organisation and effects of VET programmes;
- (b) evaluations should be planned to provide information and make recommendations on interventions to improve programmes.

⁽⁴⁾ There are, however, some differences between these cases, in particular as regards the first and the last point of the IFKA programme objectives. Furthermore, the duration is 8-14 days, as opposed to 3-5 days. The duration may also have an impact on the objectives, which are: acquisition of additional specialist or vocational skills; improvement of intercultural skills; promotion of international mobility and innovative capacities; creation of national and international networks; transfer of modern system components; international cooperation in education and trade; international marketing of German continuing training products.

⁽⁵⁾ TIMSS: third international mathematics and science study; PISA: programme for international student assessment; IALS: international adult literacy survey; DESI: *Deutsch-Englisch-Schülerleistungen-International* (German-English international student performance).

Evaluation projects are linked to quality. The question of what precisely constitutes quality is, however, often hard to answer. Admittedly, an evaluation project takes its lead from aims and appraisal criteria put forward by those involved but these must first be ascertained, exposed and, if need be, argued over at an early stage. Quality is based on normative criteria: the prescribed functions of training processes, learning goals or guiding principles of 'good practice'. Such criteria may undergo various forms of legitimisation: they may be determined by social norms, implicit or explicit political or administrative requirements, pedagogical considerations, the motives of the programme organisers or users' expectations.

Quality criteria should be made accessible to rational discussion through operational practices, rooted in theory and empirical testing of viability and implementation requirements. Training processes have multiple aims distributed simultaneously across various levels (individual, institutional and system goals) and may be variously conceptualised and weighted according to the perspectives of the different groups involved. Possibly, a given aim may be realised by various means and very different goals with varying results may be attained by the same method. This generates conflicts of interest when an option has a negative effect on one aim and a positive on another. Multiple criteria, multiple levels and multiple perspectives call for complex theoretical approaches, empirical designs and research methods. It is, therefore, worthwhile examining not only attainment of goals, but also the aims that are explicitly and implicitly set.

1.3.2. Criteria for programme evaluation

To judge the efficacy of international and EU programme implementation and impact, we need to examine current evaluation criteria. This is not easily achieved as different attitudes towards evaluation research and practice have developed. To ascertain which criteria are pertinent to evaluation projects, discussion needs to be focused on the relationship between research and evaluation.

Evaluation has greatly increased in standing in the last decade. Attitudes to the relationship

between evaluation and research have shifted dramatically. A positivist and ideals-based conception of research used to dominate basic social science research. This advocated the merit criteria of reliability, validity, objectivity and internal validity and aimed to draw conclusions on causal relationships between variables. These merit criteria, although not explicitly canonised, still play a *de facto* role today. They have, however, only rarely been attained in evaluation projects resulting in a split between research and evaluation (e.g. Wottawa and Thierau, 1998). In the context of evaluation, the matter of standards now receives more thought, indicating the specific problems associated with the attempt to comply with traditional research standards. The rise of constructivist approaches, combined with action research considerations, has led to an emphasis on the merit criteria of communication, intervention, transparency and relevance in evaluation practice. The distance between the evaluator and the matter evaluated, formerly considered essential to objectivity, was abandoned and evaluators themselves have become tools in the evaluation process. Construction, optimisation and legitimisation of programmes and measures have become key evaluation aims and the presentation of alternatives has been stressed (Stockmann, 2000; Widmer, 2000).

In the current debate a rapprochement between the different trends now seems apparent. The division has given way to a combination of evaluation and research which stresses their respective characteristics. This synthesis is apparent in the standards of the Joint committee on standards for educational evaluation (JCSEE, 1994; 2000) ⁽⁶⁾. Although the JCSEE evaluation proposals were developed for broader, national teaching programmes, most are equally applicable to the evaluation of local programmes and to fields outside education. In their analysis (in this volume) of 'ethical and normative standards for evaluation practices', Beywl and Speer conclude that the standards are interculturally transferable and applicable to VET.

The standards for evaluation of the JCSEE postulate that evaluations should be:

⁽⁶⁾ The German-speaking world has witnessed publication of a series of works that make reference to the JCSEE standards: Hager et al. (2000); Posch and Altrichter (1997); Stockmann (2000a); Thonhauser and Patry (1999) and DeGEval (2002).

- (a) useful: does a given evaluation deliver practical information on content, timing and preparation?
- (b) feasible and realistically, diplomatically and economically capable of being implemented: is the evaluation planned and conducted in accordance with real conditions and economic factors?
- (c) handled legally and ethically: does the evaluation guarantee the protection of individual rights?
- (d) accurate: are the results 'technically appropriate' and formulated using 'clean' methods according to recognised merit criteria? (Stockmann, 2000a).

These four desirable evaluation characteristics define groups of more specific standards. Another project of the third report on VET in Europe will focus on these JCSEE standards in more detail. For our purposes it should be noted that the standards serve:

- (a) as a basis for critical reflection on one's own activities;
- (b) as an aid in the various phases of evaluation;
- (c) as a basis with which to justify one's own methods in public;
- (d) as information to stakeholders on what may be expected from an evaluation and its report.

Equally, it is apparent that not all JCSEE standards may be entirely fulfilled. Often it has even been the case that in practice only accuracy standards have received appropriate attention (Schiffler and Hübner, 2000). This does not mean that norms in other categories are ignored, but their inclusion is often only implicit and they are seldom reflected in the discussion of the evaluation. Implicit agreement on certain essential principles is reflected in the similar formulations of standards by different institutions.

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. In deriving merit criteria, this group of JCSEE standards is most likely to profit from research methods on quality assurance.

The thesis can, therefore, be put forward that research and evaluation are not based on

different fundamental principles of empirical research, merely that the methodological details differ. Existing sociological methods must simply be applied consistently to evaluation projects. Consequently we largely agree with Rost (2000) and Rossi et al. (1988).

What are the peculiarities of programme evaluation to which sociological methods must be adapted?

- (a) The point of departure for any research is a question which should ideally be formulated as a hypothesis. Evaluation, however, does not examine a theoretical question, but investigates a programme as a product. The product is the result of a development process that is initially conducted without regard to its capacity for being evaluated. The product is, therefore, already available and is examined for effectiveness, generally in theoretical terms. This is even true of formative evaluations, in which the product is changed and improved but where the changes serve optimisation rather than adaptation to the methodological demands of the rest of the research process. In evaluation it is therefore appropriate to work 'backwards' and to incorporate theory into the evaluation process structure.
- (b) A programme is usually characterised by a high degree of complexity. The analytical process of isolation or systematic manipulation of variables is thus questionable. The isolated contemplation of individual components and subsequent investigation of their interaction does not do justice to the complexity of the matter to be evaluated. In such cases, methods such as path analysis, which can track complexity, are preferable ⁽⁷⁾.
- (c) Maintaining disruptive factors at a constant level is rarely practicable, since local conditions are so variable. Particular attention must therefore be given to such factors which might influence data.
- (d) Constructs and variables are already in operation. Theoretical aspects of measurement must therefore also be constructed 'backwards'.
- (e) The principle of randomisation does not usually apply. The representativeness of the

⁽⁷⁾ See remarks on IFKA in Section 6.2.

sample population and the implications of the results are therefore limited. The range of applicability of the results of an evaluation must therefore be borne in mind.

Despite these differences, structural similarities exist between basic research and evaluation. Although the methodological formation may differ in some respects, the basic intentions remain comparable.

With regard to the three different phases put forward by Rossi et al. we can formulate the structural similarities between research and evaluation as:

- (a) conceptualisation phase: corresponding to generating and incorporating the hypothesis in basic research;
- (b) implementation phase: corresponding to making variables operational, commissioning a design, ascertaining independent variables, and the 'manipulation check', i.e. the test of the extent to which a desired treatment is realised and can affect participants;
- (c) impact research: corresponding to the various measures to ascertain impact in basic research, e.g. capture of dependent variables, controlling for third variables and statistical checks and calculation of effect size.

In the light of these structural similarities between research and evaluation, compliance with standards does not mean using particular methods but showing that suitable methods have been applied in each phase, as follows.

Conceptualisation phase

Every activity to be evaluated must have an underlying concept. This may have to be reconstructed. Empirical practice without theory is inadmissible. Evaluation must also be characterised by a 'theory-driven approach' (Chen and Rossi, 1983).

A scientific evaluation should include a theoretical explanation of relevant processes which are to be given due consideration during planning of the study and interpretation of the results.

A report on a programme's impact is of less value if the results are not based on a theoretical construct. Because of the frequent lack of systematic variability and the want of representative population samples, generalisation is only possible via theoretical models.

The criteria for the conceptualisation phase are therefore:

(a) Conceptualisation

The theoretical bases of measures or programmes to be evaluated must be reconstructed in line with these considerations:

- (i) What is the assumed impact?
- (ii) What criteria are to be met?
- (iii) What is to be improved?
- (iv) Do the goals target specific or general effects?
- (v) Are goals precisely defined?
- (vi) Do the goals fulfil European or national needs?
- (vii) Does the choice of target groups comply with programme goals?
- (viii) What does the postulated process that triggers the impact consist of?
- (ix) In what areas do effects come about?
- (x) Is the concept internally consistent?
- (xi) Can means of optimising the programme be derived from the structure of evaluation?

(b) Questions and hypotheses

Conceptual bases must be linked to the measures or programmes to be evaluated in the form of research questions and hypotheses which may be empirically verified with the help of a realisable evaluation design.

Implementation phase

Implementation is more important in evaluation than treatment control in basic research.

In this phase the stress is not yet on the impact but on operationalising the active variables.

The task of an implementation study is to ascertain the intensity of an activity and to identify possible disruptive variables which are to be monitored in an impact study.

Checks on the implementation of an activity should investigate the following aspects:

- (a) to what extent the realisation of an activity corresponds to its concept,
- (b) how great is the impact of the influences under investigation,
- (c) whether other influences favour or hinder potential impact.

Impact research phase

This phase, often considered the true task of evaluation, is the one most closely associated with basic research.

To consolidate impact, methodological principles should be adapted to the peculiarities and limitations of evaluation studies.

Treatment of the transferability of results and the causal interpretation of effects are of primary concern.

(a) Transferability of results

If no random sample is feasible, the sample investigated should be comprehensively described in order to clarify general validity. If necessary, selected groups should be investigated in detail, to support generalisations and specific findings.

(b) Causal interpretations

Since randomisation enables some impact control of 'moderator' variables, any evaluation which does not permit random population sampling should incorporate as many potential moderator variables as possible to ensure causal interpretations.

(c) Control groups

Control group design is an indispensable procedure in basic research. In evaluation studies, however, control groups are often hard to create for programme reasons. They are unrealistic, unavailable or difficult to define. The actual purpose of control group design is the creation of a comparative group to highlight one variation of the independent variable. This is a prerequisite for contextual validity. This aim of creating a comparative standard by which impact can be ascertained may also be achieved by:

- (i) dosage variation, e.g. duration or intensity of the measure to which participants are exposed;
- (ii) facet design: comparison of populations exposed to different component combinations of a given measure. This permits statements to be made on the size of impact, depending on the effects of various components and their interaction;
- (iii) comparison of populations which differ in theoretically defined characteristics from the evaluation population. In general, individual programmes and their projects create different patterns of initial requirements and implementation procedures due to the given variables. Quasi-experimental designs may be derived by induction. The contrast method allows system-

atic comparison of members of different configurations to arrive at explanatory hypotheses. By the further inclusion of project participants, the plausibility of the interim hypotheses may be further tested and modified as need be until a satisfactory 'saturation' has been reached. This is a financially viable alternative to more complicated large-scale projects that are essentially quasi-experimental from the outset.

In the context of evaluation of microeconomics, Hujer, Caliendo and Radi refer in this report to the debate on the varying levels of acceptance accorded to non-experimental methods. They advocate methods capable of coping with the problems generated by non-experimentally-won data, such as the before-after estimator, the cross section estimator, the matching estimator, the difference-in-difference estimator and the duration model approach. All these processes attempt to estimate an appropriate comparative standard within, or on the basis of the relevant data set.

Evaluations should, therefore, draw on variations in impact factors and programme variants to strengthen the plausibility and general validity of findings.

(a) Statistical checks

If potential disruptive and moderator variables have been included, these can be incorporated in the statistical analysis. With the help of partial correlation, covariant analysis, step-wise regression and linear structure comparison models, effects can be statistically controlled or estimated.

(b) Effect size

Particularly in evaluation, although not exclusively, mention of statistical significance is either superfluous because of sample size or self-evident. Nonetheless, the statistical significance of ascertained differences and relationships should be cited as evidence of the impact of an activity. Effect size is particularly important. Effect size describes the size of the observed effect in proportional variance, correlation, standardised mean deviation or in survey tool measurement units. These are instruments capable of quantifying the practical usefulness of measures (Hager, 2000; Rost, 2000; Sedlmeier, 1996).

The above criteria make high demands on programme evaluation. We cannot expect individual programmes to meet all the above criteria. It is more a matter of individual

programmes setting their own priorities. The following case studies should be examined to see what aims they strive toward and what criteria they fulfil.

2. Case study 1: the LdV programme (mobility)

2.1. General framework

The LdV action programme was launched in 1994. As a central vocational training support programme, it was partially or fully responsible for triggering several initiatives within this sector (Comett, Eurotecnet, FORCE, PETRA and Lingua).

The programme is now in its second phase (2000-06). Its priorities were established in a Council Decision of 26 April 1999. They focus on further developing the quality, innovativeness and European dimension of vocational training systems by promoting transnational cooperation.

The countries eligible to participate in the programme are:

- (a) the associated countries of Central and Eastern Europe (CEECs), under the European Treaty and additional protocols, already or yet to be concluded, on the participation of these countries in Community programmes;
- (b) Cyprus (its participation is financed by additional funds under provisions also applicable to countries in the European Free Trade Area [EFTA] which are members of the European Economic Area [EEA]);
- (c) Turkey and Malta, under procedures yet to be agreed with those countries.

The measures in the second phase of LdV can be divided into five separate categories:

- (a) mobility,
- (b) pilot projects, including theme-based initiatives,
- (c) language skills,
- (d) transnational networks,
- (e) comparative materials.

The following comments refer only to the area of mobility, which covers the type of case studies presented here.

Extract from the general guidelines for applicants in the area of mobility ⁽⁸⁾:

‘Proposals submitted under this heading must relate to transnational mobility actions for people in training, more especially young people under-

going training or endeavouring to enter the labour market, young employed workers or job seekers, recent graduates, students registered in higher education establishments. They may also be aimed at trainers or human resource managers and training scheme managers, at language specialists and at social partners.

The basic point here is to strengthen the European dimension of initial and continuing vocational training, to encourage people to gain experience in activities involving theory and practice, particularly work-linked training, to develop language skills and transnational contacts and exchanges of good practices for trainers and human resource managers.

This type of measure is also useful in consolidating transnational cooperation in a broad sense, bringing in all actors on the vocational training scene. It is also a good way of strengthening the links between working life and training.’

The actual implementation takes place within three main areas of action, depending on the groups of beneficiaries involved.

- (a) Transnational placement projects:
 - (i) for people in initial vocational training: in a vocational training institution or enterprise. Objective: promoting the acquisition of complementary vocational skills in another Member State to improve access to the labour market;
 - (ii) for students: in an enterprise. Objective: promoting entry to the labour market. This group of people can also help enterprises gain transnational experience and promote the application of technology transfer;
 - (iii) for young employees and recent graduates: in a vocational training institution or enterprise.
- (b) Transnational exchange projects: these are intended for trainers, careers advisors and educational guidance counsellors, etc.
- (c) Study visits: for vocational training decision-makers, including the social partners. Objective: to promote the transfer of technological

⁽⁸⁾ Not to be confused with the CSV programme for vocational training decision-makers.

innovations in SMEs and to provide continuing training for company managers to keep pace with technological and organisational changes; to prepare transnational vocational training campaigns.

(i) Placements usually last:

- from 3 weeks to 9 months for those undergoing initial vocational training
- from 3 to 12 months for students
- from 2 to 12 months for young employees and recent graduates

(ii) Exchanges last between 1 and 6 weeks.

The exchange and placement mobility projects are subject to national procedure A (a single-step procedure). In the case of Germany, applications are submitted either to the Education for Europe National Agency at BIBB or to the relevant implementation centre (InWent, DAAD, Bundesanstalt für Arbeit [BA] or ZAV). The implementation centre checks applications and prepares the choice of projects for the selection committee. It also advises on all questions relating to project application, implementation and supervision.

2.2. Programme objectives

The programme was primarily introduced to promote a 'Europe of knowledge' and to consolidate European cooperation in general and VET. The Council Decision of 6 December 1994 on an action programme to implement a European Community vocational training policy formed the basis for the formulation of objectives. That document lists the common goals:

- (a) improving the quality and the innovative capacity of Member States' vocational training systems and measures;
- (b) developing the European dimension in vocational training and guidance;
- (c) promoting lifelong vocational training and supporting associated policies;
- (d) encouraging vocational training measures for unskilled adults;
- (e) enhancing the status and attractiveness of VET, and fostering equivalence between academic and vocational qualifications;
- (f) promoting vocational training for young people and preparing them for adult and working life;

- (g) encouraging vocational training measures for disadvantaged young people lacking adequate training;
- (h) promoting equality of access to initial and continuing training for disadvantaged persons;
- (i) promoting equality of opportunity as regards access for women and men and their effective participation in vocational training;
- (j) promoting equality of opportunity as regards access to and effective participation in, vocational training for migrant workers and their children, and for disabled people;
- (k) promoting cooperation in respect of skill requirements and training needs and encouraging the acquisition and transparency of qualifications and an understanding of the key skills relevant to technological development;
- (l) promoting vocational training in the light of the results of technological research and development programmes;
- (m) promoting the gradual development of an open European vocational training and qualifications area;
- (n) supporting activities aimed at developing linguistic skills as part of vocational training measures;
- (o) promoting the development of vocational guidance facilities;
- (p) fostering the development of methods of self-training in the workplace and of open and distance learning and training;
- (q) encouraging the development and integration of key skills in vocational training measures;
- (r) giving all young people in the Community the opportunity to undergo one or, if possible, two or more years of initial vocational training on top of their compulsory education.

The second phase of the programme follows three general objectives:

- (a) improving the skills and competences of people, especially young people, in initial vocational training at all levels. This may be achieved *inter alia* through work-linked vocational training and apprenticeship, with a view to facilitating vocational integration and reintegration;
- (b) improving the quality of, and access to, continuing vocational training and the lifelong acquisition of skills and competences, with a view to increasing and developing adapt-

ability, particularly in order to consolidate technological and organisational change;

- (c) promoting and reinforcing the contribution of vocational training to the process of innovation, with a view to improving competitiveness and entrepreneurship, and also considering new employment possibilities. Special attention shall be paid in this respect to fostering cooperation between vocational training institutions, including universities, and companies, particularly SMEs.

The objectives of this programme shall be achieved by means of the following measures:

- (a) support for the transnational mobility of people undergoing vocational training in Europe and of those responsible for training;
- (b) support for pilot projects based on transnational partnerships designed to develop innovation and improve quality in vocational training;
- (c) promotion of language skills, including less widely used and taught languages and understanding of different cultures in the context of vocational training;
- (d) support for the development of transnational cooperation networks facilitating the exchange of experience and good practice;
- (e) development and updating of Community reference material and comparative data.

In implementing these actions, the Decision provides for particular support for transnational actions to promote and use information and communication technologies in vocational training.

The programme objectives have the following priorities:

- (a) enhancing employability by making VET systems more efficient in Europe;
- (b) promoting partnership between training institutions at all levels, the business sector and the social partners;
- (c) facilitating the integration of disadvantaged persons in the labour market and countering discrimination;
- (d) encouraging companies to invest in human resources as a business strategy;
- (e) exploiting the potential of information and communication technologies;
- (f) improving the transparency of qualifications.

2.3. Evaluation of the overall programme, individual programmes, particular aspects, etc.

2.3.1. Overall programme

We divide the evaluation of LdV mobility measures into two separate areas: implementation of the programme in the EU and implementation of the programme in individual countries (e.g. Germany).

2.3.1.1. Implementation of the programme in the EU

Objectives of evaluation of the overall LdV programme, focusing on the effects for the EU:

- (a) comparing the programme's priorities and the quantitative data collected;
- (b) identifying differences in the individual countries involved in the programme;
- (c) determining trends in European vocational training systems;
- (d) assessing vocational mobility promotion tools practised throughout Europe (e.g. Europass);
- (e) presenting transnational difficulties regarding the programme context (application procedure, financing, etc.).

Evaluation reports:

- (a) Commission interim report of 23 July 1997 on the implementation of the LdV programme [COM (97) 399 final, not published in the Official Journal]

The report is based on both the results of an objective external evaluation and the participating countries' reports on the implementation and impact of the LdV programme.

Extracts:

Some general trends emerging from the national reports are as follows:

- (i) concern for employment and employability is becoming increasingly pronounced in vocational training policies;
- (ii) all the reforms undertaken are aimed at improving vocational training and guaranteeing its quality;
- (iii) vocational training is becoming decentralised and individualised with more responsibility on those involved at local and regional levels and the social partners.

All the Member States stress the importance of measures geared to improving the prospects for social and occupational integration of vulnerable target groups such as unskilled young people, young unemployed people, the long-term unemployed, people with disabilities and immigrants.

The Council Decision estimated that the financial resources required at Community level to implement the LdV programme would be ECU 620 million for the period from 1995 to 1999. The budgetary authority allocated an overall sum of ECU 139.5 million in 1995 and ECU 154.4 million in 1996. In 1995 and 1996, providing funding for 1 542 projects totalling ECU 201.8 million, the LdV programme enabled more than 50 000 persons, including 22 000 young people undergoing initial training, 10 000 young workers and 13 000 higher education students and graduates, to benefit from a European mobility grant for periods of guidance and training in the Member States and participating countries.

The Commission portrays the impact of the programme as apparent on different fronts:

- (i) the pilot projects are part of the development of the Member States' vocational training systems and arrangements; the key players involved in training see the programme as an investment priority;
- (ii) by proposing innovative solutions with regard to new qualifications, occupational profiles and means of access, rapprochement of schools and enterprises and new pathways for occupational integration and stabilisation, the programme has a significant impact vis-à-vis innovation (European laboratory for innovation);
- (iii) the programme is often perceived as an opportunity to amplify an existing project by fostering transnationalism and multi-partner cooperation;
 - the programme makes a major contribution to the development of transnational mobility;
 - the impact is even greater when the project succeeds in developing genuine training strategies.

The impact of the programme will no doubt take time to materialise, but the cooperation which has been initiated and the efforts which

are being made to respond to the continuing changes seem promising. By encouraging the parties concerned to work together, LdV has made a significant contribution to harmonising vocational training.

One of the shortcomings of the programme is the poor dissemination of project findings. The European Commission is currently paying particular attention to measures aimed at improving and disseminating findings at European level. The complexity of the programme itself and the cumbersome administrative procedures constitute an additional drawback. The conclusions drawn from experiences during the first two calls for proposals from the national reports and the external evaluation report can only help to improve the effectiveness of the programme.

- (b) Commission report of 22 December 2000, on the implementation of the first phase of the Community action programme LdV [COM (2000) 863 final].

The report is based on:

- (i) an external evaluation report by the consultancy firm Deloitte & Touche;
- (ii) national reports from the Member States;
- (iii) final reports from the social partners.

LdV has been invaluable in the promotion of transnational initiatives and the internationalisation of best practices in vocational training in the quality and content of learning, innovation and the addition of a European dimension. Its particular strengths have been the enhancement of mobility and employability of the participants in the projects. It has thereby benefited not only the participants from Member States and EFTA-EEA countries, but also those from the applicant countries where LdV projects have assisted in the preparation of national training systems for accession. These successes are the foundation upon which the second phase of the programme is being built.

However, not only the strengths but also the weaknesses of the first phase must be addressed in order to create a solid foundation for the second phase of the programme. The problems experienced during implementation of the first phase relate not only to the unsatisfactory performance of the contracted technical assistance office, but also to

complexities within management at the central level, and the lack of reciprocity with other training-related programmes. The European Commission has ensured that the lessons learned from these experiences are put into practice in the second phase, through the simplification of procedures and further decentralisation of the programme's management. It also anticipates a clear dissemination strategy for the products and outcomes of the many transnational pilot projects in the first phase.

LdV stands to become a key instrument in the implementation of lifelong learning strategies which offer synergies between the European policies for education, training and employment. Furthermore, the second phase will seek to involve some players more fully in transnational vocational training, particularly the social partners and SMEs. This strategy reflects the European Commission's policy of promoting reciprocity between related programmes, which receives a strong emphasis in the second phase.

2.3.1.2. *Implementation of the programme in Germany*

Objectives of evaluation of the LdV programme, focusing on the effects for Germany:

- (a) identifying the effects and deficits of LdV in Germany;
- (b) comparing programme objectives with actual effects;
- (c) summarising the findings from evaluations of individual measures or from evaluations of programme parts conducted by implementation centres (DAAD, InWent, ZAV);
- (d) assessing the work of implementation centres;
- (e) improving efficiency of LdV measures in Germany;
- (f) submitting proposals to the national agencies and the implementation centres.

Evaluation practice:

The *Institut für Wirtschaft- und Sozialforschung* (Economic and Social Research Institute, WSF) in Kerpen evaluated the first phase (1995-99) of the LdV programme. It is also engaged in the evaluation of the second phase. The methodological approach has now changed. No time was available to conduct thorough surveys for the evaluation of LdV I because the programme was approved at the last minute. The European

Commission has paid more attention to the question of evaluation in LdV II. Another consequence is that evaluations have been commissioned at an earlier stage, which has left more time to develop the methodology.

Methodological approach of the evaluation:

- (a) LdV I, steps in the investigation (Friedrich and Schumacher, 2000)

Performing a brief overview analysis which examines the situation and the current trends and problems of the national education system. This analysis provides a frame of reference for assessing the appropriateness of the measures implemented and of their emphases.

Appraising implementation reports, compendia, application documents, accountability reports from national coordination centres and projects, etc.

Drawing up tables and overviews summarising the material and financial progress of each part of the LdV I programme.

Systematically recording the projects supported. A survey of around 600 projects was conducted. The sample reflected the distribution of projects across the individual project fields. The survey identified the central goals, the dissemination measures undertaken and how the project findings could be utilised. Over 200 projects responded to the survey. The time restrictions mentioned above made it impossible to increase the response rate by sending out reminders. The data of those projects which provided answers were weighted to relate to the total number of projects supported. The findings thus provide a representative picture of all measures supported.

Analysing the strengths and weaknesses of the programme by consulting the players involved and independent education experts. The latter can be divided into three categories: members of the national LdV committee and subcommittees, members of project advisory boards or similar project support bodies and project partners. Around 80 independent experts provided information on the strengths and weaknesses of LdV.

In addition, programme coordinators shared their experiences through structured interviews.

(b) LdV II, method

In contrast to the evaluation of LdV I, which had to be completed within just three months, it was possible to issue a long-term contract for the evaluation of LdV II. This made it possible to adopt a multi-stage methodological approach, something that is particularly important for data on the sustainability of mobility projects.

Examination of the LdV programme draws a fundamental distinction between mobility measures and 'innovative projects'. The latter have particularly required and still require, close cooperation with the National Agency. Since the evaluation project was awarded to WSF, several preparatory meetings have taken place (as many as six per year) with the National Agency and the implementation centres.

The evaluation of mobility measures is primarily based on participant questionnaires. In the first stage at the end of 2001, WSF sent questionnaires to around 1 000 interviewees three to six months after they had participated in the LdV exchange programme. This first group consisted of school pupils, young people in initial vocational training, young employees and multipliers. Because of the specific length of the programmes students were not involved in the first year.

The questionnaire was drafted in close cooperation with the implementation centres, which also use the WSF's questionnaire model for their surveys.

In the second stage at the end of 2002, 1 000 new participants (this time including the students) were surveyed, again three to six months after completing an exchange programme. The 600 members of the first group who had completed questionnaires the first time were also questioned again, so that the respondents now totalled 1 600. In the third year, 1 000 new participants and the people from the first and second groups will be surveyed.

In this way the evaluators hope to gain a detailed picture of the effects of exchange programmes by the end of the study.

Evaluation report:

Friedrich and Schumacher: *Evaluierung der Umsetzung des Programmes LdV in Deutschland*

1995-1999 (Evaluation of implementation of the LdV programme in Germany from 1995 to 1999). An enquiry commissioned by the Federal Ministry of Education and Research. Kerpen: January 2000.

(a) The key findings were:

- (i) the European Commission had a major influence on the choice of projects;
- (ii) in Germany coordination was divided among seven centres; BIBB had overall responsibility;
- (iii) 56 % of participants were women;
- (iv) social partners, experts and authorities were adequately represented on committees;
- (v) application forms were too long and complicated;
- (vi) online project applications did not function properly.

(b) Extracts from the report relating to mobility projects:

About 2 500 LdV projects have been implemented in Germany to date. Pilot projects, multiplier projects, surveys and analyses account for approximately 13.7 % (340) of all projects since LdV started. Placements and exchange programmes account for 86.3 % (around 2 100 projects).

LdV exchange measures have involved approximately 30 000 people. The figures showing the percentages of different types of participants in LdV exchange and placement programmes are: trainees 42 %, young employees 17 %, students/recent graduates 30 %, trainers, etc. 10 %. The programmes have only involved about one thousandth of the total population of all these groups, except for trainers. Consequently the programme has had no remarkable quantitative effects as might be expected given the low investment. However, we should mention that national bodies provide few opportunities for comparable support programmes, so that LdV has at least represented a useful enhancement.

Project organisers did not conduct systematic impact analyses of these placement and exchange projects. Furthermore, insufficient time and resources were available for the intended ex-post evaluation. This was due to budget restrictions imposed by the Commis-

sion. But analyses of this nature are crucial if we are to identify the conditions under which these measures are particularly successful. Nevertheless, the findings of expert surveys during ex-post evaluations and isolated investigations undertaken independently by the relevant specialist coordination centres suggest that these measures are the most effective of all support programmes.

Target-group specific exchange programmes:

Most projects have been aimed at young people undergoing initial vocational training. Since 1995, 900 measures have been carried out under heading I.1.2.a. This represents 36.5 % of all measures implemented in Germany and 42.3 % of all exchanges. The longer-term exchanges offered to the same target group have played a relatively insignificant role, totalling 64, i.e. 2.6 % of all projects, 3 % of all exchange projects. The measures gave young people a chance to gain additional vocational skills.

Exchanges for young employees have consisted of continuing training schemes and work placements abroad. Continuing training courses account for 6.2 % and placements abroad for 15.3 % of all projects supported since 1995.

Transnational placement programmes for students or recent graduates account for 3.5 % of all projects. The low number of student placements should not detract from their significance. This type of project is characterised by the large number of partners involved and the high number of participants (almost 9 000 since 1995). The projects have contained bundles of measures involving students at German universities going abroad and foreign students assigned to German companies.

All programme areas have featured projects encouraging exchange of trainers. However, only those incorporated in programme area I are significant in both number of projects and number of participants, representing 21.4 % of all projects. Additional exchange projects have included improvement of language skills (a total of six projects) and schemes for educational decision-makers (although no precise figures are available).

The survey of project providers revealed the objectives of exchanges in programme area I, which focused on initial vocational training, to be

increasing specialist competence, presenting new techniques and introducing different cultures. Each of these goals received 75 points on a scale of 0 to 100. Far less significant were boosting company competitiveness and, astonishingly, preparing young people to embark on careers. Each received 23 points. These assessments clearly demonstrate the priorities of short-term exchanges for young people in initial vocational training, who overwhelmingly dominate this programme area.

In programme area II the main priority was to respond to the demand for European mobility (100 points), followed by improving specialist competence (98 points) and acquiring language skills (93 points). Promoting the appeal of vocational training and general exchange of information and experience had little or no relevance.

Impact of LdV further training measures:

(a) Participants

There has been no systematic investigation of the impact of LdV placement and exchange projects at participant level (e.g. via surveys). The intended analysis of this kind could not be realised as part of the ex-post evaluation because of time and cost constraints.

However we can affirm that, while the exchange programmes do not fully fill a gap in the German education system, they have made a significant contribution towards doing so. The findings from the expert survey hint at the same conclusion. The experts believed that of all programme components, placements and exchanges had had the most impact. These were awarded the very high score of 81 on a scale of 0 to 100.

Since no participant impact indicators are available, the following programme impact analyses are mainly based on providers' and experts' comments. It should be borne in mind that providers' remarks are often biased since they tend to overestimate their own projects, at least as far as scale of impact is concerned.

(b) Projects

The impact of placement and exchange programmes on project providers is similar to that of the pilot schemes. The programmes also achieved a distinct improvement in competence. Characteristic of these projects is that they directly target participants, yielding the following effects:

- (i) a more positive attitude towards European mobility among participants (78 points);
 - (ii) improved key skills with particular relevance for the European internal market (73 points);
 - (iii) increased transnational experience (72 points);
 - (iv) improved language training (71 points).
- (c) Programme level
- The impact of the exchanges seems to have been extraordinarily high (81 index points) but one must remember that the quantitative scope was limited from the outset due to the low level of funding available for measures of this nature.
- The programme also helped to reinforce the important cooperation between schools, training institutions and enterprises (68 points), a crucial area of practical initial and continuing vocational training.

2.3.2. Evaluation of individual measures and particular aspects, in Germany and elsewhere

Objectives of the evaluation of measures:

- (a) quality assurance,
- (b) provision of support and advice to projects,
- (c) feedback for evaluators' own work,
- (d) comparison of project objectives with actual impact,
- (e) increased efficiency of individual projects,
- (f) often simply a source of information for the implementation of the programme in Germany,
- (g) no systematic assessment of data (much of this occurring directly).

2.3.2.1. LdV evaluation practice at the ZAV

In Germany applications for LdV mobility projects are handled and approved by different offices, depending on which target group applicants belong to. These are:

The following statements refer to ZAV's particular area of activity, providing work placements abroad for young employees.

This is a heterogeneous target group, since people of all ages can participate in the measures. Because ZAV is an institution of the German Federal Labour Office, its implementation of LdV projects usually focuses on the labour market and worker employability. This distinguishes the organisation from the other two implementation centres, InWent and DAAD, which concentrate more on education, training and pedagogical issues.

Two different categories of evaluation exist:

- (a) Category 1: LdV programme evaluations in Germany
- WSF in Kerpen, which carried out the evaluation of the first phase, is also evaluating the second phase of the LdV programme. Evaluation has involved surveys of the implementation centres. A report on the first phase of the LdV programme is available. To date, however, it has received limited attention.

Table 1: Applications for LdV mobility projects: target groups and implementation authority

Target group	Duration	Implementation authority
Young people undergoing initial vocational training	short: 3-12 weeks long: 13 weeks – 9 months	InWent
Young employees: • in work placements abroad • in vocational (school-based) continuing training	8 weeks – 12 months	• ZAV • InWent
Students and graduates on work placements (from university/college)	3-12 months (students) 2-12 months (graduates)	DAAD
Vocational training decision-makers: • Trainers • Academic and business staff exchanges	1-6 weeks	• InWent • DAAD

- (b) Category 2: evaluation of the measures supervised by the relevant implementation centres (ZAV; DAAD; InWent)

ZAV pays greatest attention to the evaluation of measures. During the first phase of LdV it conducted its own surveys of participants who had spent between three and six months working abroad. The response rate was between 60 and 70 %. The focus was on collecting data for labour-market statistics.

European Commission regulations stipulate that implementation centres must evaluate and monitor LdV II projects. ZAV is therefore currently conducting a participant survey based on experiences gleaned during LdV I. The survey has been adapted to the requirements of InWent and DAAD so that it can be used for programme evaluation (category 1). In contrast to earlier studies, which served purely statistical purposes (e.g. reintegration into the labour market, levels of satisfaction with ZAV) and were therefore anonymous, the latest participant surveys refer to specific individual measures.

The three implementation centres are planning a joint strategy for conducting future project evaluation.

- (a) Monitoring

Continuing project evaluation or monitoring will increase in significance. InWent has developed guidelines for project visits in Germany and abroad. They are intended to facilitate systematic assessment of these visits. Projects received spontaneous visits during the first phase but no cross-project assessment of findings took place.

- (b) Providers' interim and final reports, participants' reports

Providers' and participants' reports are still the most important element of LdV project

evaluations. The implementation centres exploit the reporting obligation prescribed by the European Commission. Participants and providers have submitted to ZAV interim and final reports on approximately 50 approved projects (out of 60-80 applications). The reports have a uniform structure.

Before the final project payment can be made, the success of a project is assessed by comparing the application with the final report. The results of this assessment are not recorded.

Participants' reports are analysed with the help of a table which records their assessment of the most important project phases.

The implementation centre records any conspicuous entries and discusses them directly with the providers of the relevant programme. No further assessment of the data from participant reports is planned.

- (c) Criticism and recommendations

More thorough recording of the attitudes and opinions of enterprises providing placements would be desirable. Since ZAV focuses its activities on the labour market and on employability, the enterprises involved in LdV clearly have a crucial role.

A sustainability study should certainly be conducted but should be assigned to an outside body because of implementing centres' limited time and financial resources. This would have the advantage of alleviating providers' fears of being disregarded when new project contracts are awarded.

Another way to increase the range of quality assurance measures would be to integrate individual evaluation steps into the project, i.e. project providers would conduct sustainability studies themselves.

Table 2: **Raster for assessment of the most important project phases of LdV mobility projects from the point of the participants**

	Good	Average	Unsatisfactory	Reasons
Information/preparation/organisation				
Language course				
Accommodation/catering				
Accompanying programme				
Support offered in host country				
Placement				

2.3.2.2. LdV evaluation practice at InWent

(a) Conference, annual general meeting

The LdV annual general meeting is regarded as the most important link in the programme and project chain. It is seen as a service for providers, permitting dialogue between the providers of individual measures and the implementation organisations. It offers 250-300 project providers a platform to discuss any problems that have emerged and any difficulties that may be foreseen. Since 2002, the annual general meeting has been jointly staged by the three implementation centres exclusively for LdV projects. The advantage of this over earlier meetings, which also included Socrates providers, is that discussions are focused on the matter in hand. The details given in project applications on the intended evaluation process are one of the selection criteria for project sponsorship.

(b) Providers' interim and final reports, participants' reports

The appraisal of providers' and participants' reports has changed from LdV I to LdV II. In the first phase, only around 10 % of providers' reports (final reports and interim reports if projects lasted over 18 months) were closely examined. Even participants' reports were only spot-checked.

InWent is aiming for a higher standard of report evaluation for the second phase of the programme so that 20-30 % of providers' reports are investigated according to a standardised process. The German Craft Trades Institute in Cologne has drafted an assessment form expressly for this purpose. Participants' reports remain relatively unstructured, but more are assessed.

In the EU's view, the provider of an activity is the key partner of the implementation centre. Assessment results are submitted to the European Commission once a quarter.

(c) Monitoring

In the future, monitoring will play a key role in quality assurance. Arranging more supervisory visits is one of the European Commission's requirements for the implementation of LdV II. Onsite visits were rare in the first phase and only occurred if problems arose in the host country. At least four and usually six

annual supervisory visits abroad are planned for the second phase.

A set of guidelines for supervisory visits has been developed as methodological support for monitoring. It categorises results as follows: overview of the entire project; products/interim results/final results; dissemination strategy; work plan; partnerships; financing/budgeting of grant and results of financial audit.

In Germany 10 projects will be subjected to financial audit. The most important selection criterion for these 10 projects will be the value of the contract, followed by the target group and other pertinent data. Of these 10 projects, 5 will be chosen to be the focus of supervisory visits by InWent in the context of an assessment seminar. Standardised reports of the outcomes of these reports will be drawn up.

(d) Treatment of evaluation findings

Once a year, about three months before the application deadline, there is an opportunity to make use of the findings and experiences of previous years in designing new projects. For example, evaluation reports are used when updating Internet homepages in order to draw attention to possible changes, e.g. accountancy procedures, duration of placements, etc. Various InWent task forces focus on different issues during these three months. This has led to publication of a booklet on long-term measures.

2.4. Summary

The second phase of the LdV programme pays more attention to evaluation design for both individual measures and the programme as a whole. At the application stage itself and during subsequent monitoring of activities, questions of evaluation play a significant role in determining whether the measure receives support.

Each year the measures which fall under LdV are evaluated. This is a major task but the coordination of the individual measures could be improved by those involved in this area. Often data are collected which help implementation centres to improve the quality of the programme or its measures in the short term but which are no

longer available for subsequent, more in-depth investigations.

To date, evaluation methods have not usually appeared to take into account the stated programme objectives. Instead, the priority has seemed to be whether a project was successfully organised and whether participants did as well as possible.

A look at the most important objectives of LdV I (Section 2.2) reveals that the programme designers wanted to avoid expressing measurable goals. In principle they were hoping to gain support for vocational training throughout Europe without spelling out the opportunities for intervention offered by transnational exchange programmes.

Therefore, the evaluation reports on LdV I contain little information on the actual, verifiable impact of LdV mobility measures on vocational training in Germany. 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.

It is to be expected that the final report on the evaluation of LdV II will contain more references to goal attainment.

Core objectives of LdV II:

- (a) improving skills and competences of people, especially young people, in initial vocational training, with a view to promoting their employability at all levels;
- (b) improving quality of continuing vocational training and access to lifelong acquisition of skills and competences, with a view to increasing and developing adaptability, particularly to technological and organisational change;
- (c) promoting and reinforcing the contribution of vocational training to the process of innovation with a view to improving competitiveness and entrepreneurship and creating new employment opportunities.

Improved coordination of evaluation of projects (ZAV and InWent) and programme (WSF) will make it possible to comment about employability (ZAV) and continuing vocational training (InWent and WSF).

In general, the method adopted for the evaluation of the LdV programme and its impact is praiseworthy, since the three-year study involving several groups should permit conclusions to be drawn about objectives (a) and (b).

However, useful findings regarding objective (c) are unlikely. At the moment, evaluators do not systematically consider the comments and opinions of the enterprises involved. The only contribution the evaluation might make towards fulfilling objective (c) would be in providing a best practice model (Lauterbach, 2003, p. 108 et seq.).

More systematic assessment of providers' and participants' reports could provide additional information on the impact of mobility measures on vocational training. However, appraisals of this nature are very demanding, and implementation centres are not normally able to carry them out. One prerequisite would be the introduction of a reporting system with a standardised report model, increased assistance in writing reports and investigation of appropriate electronic assessment procedures. It would be sensible to commission third parties to carry out these tasks.

If the standards of the JCSEE (1994, 2000) and the German Evaluation Society (DeGEval) are taken as yardsticks, the value and usefulness of LdV evaluation could certainly be increased. According to DeGEval the four basic characteristics of an evaluation are utility, feasibility, fairness and accuracy. These four basic characteristics can be divided further into sub-categories. Those of particular relevance to increasing the quality of LdV evaluation are to be found under Point 1 as 'clarification of the purposes of the evaluation' (more precise definition of the purposes of the individual stages of programme and project evaluation) and 'timing of evaluation' (contracting evaluations early so that soundly based methods can be developed and sufficient time allowed for coordination with the institutions involved, implementation centres and national agencies; under Point 2 as 'appropriate procedures' (more efficient use of the reporting system, avoidance of repetition of statements of participants in reports and questionnaires) and as 'efficiency of evaluation' (sensible division of labour, e.g. drawing in third parties or providers of individual projects); and under Point 4 as 'context analysis' (supporting research into the Europeanisation of vocational training in training institutions and

enterprises) and ‘meta-evaluation’ (processing of data for further investigations).

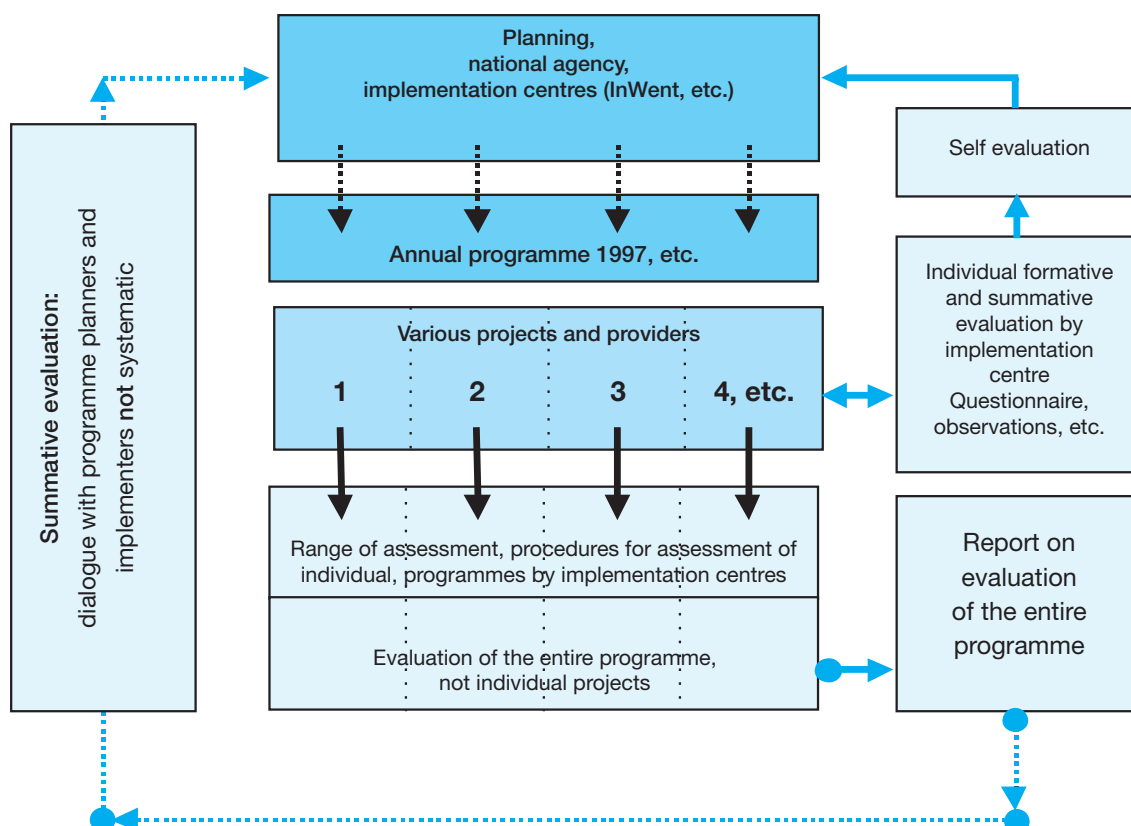
Evaluation of the practical and financial aspects of the LdV mobility programme and its individual projects should be guided by JCSEE or DeGEval standards. Our recommendation would be to take a bottom-up approach, since the findings of studies into individual projects could be incorporated into evaluation of the mobility programme and subsequently of the entire LdV programme. It would seem sensible to divide up evaluation of the programme as a whole according to project type (mobility, innovative projects, etc.) because of the diversity of measures.

Despite this prior division, cluster evaluation would be a suitable methodological approach to evaluating LdV mobility projects. The heterogeneous financial and structural frameworks of the mobility projects demand an evaluation method developed for programmes with a common purpose but whose individual projects are relatively independent and have varying strategies, contexts, etc.

Key characteristics of cluster evaluations are as follows:

- (a) cluster evaluations investigate a group of projects to identify shared themes through comparative observation;
- (b) they do not just try to understand what happens within projects, but also ask why;
- (c) they are carried out in cooperation with all those concerned. Projects managers providers, those commissioning the evaluation and evaluators themselves are all involved in the process and everyone can profit from the findings;
- (d) project managers and evaluators have a confidential relationship. Information is only released if it is relevant to the programme, but data on individual projects are never divulged. This guarantees that project coordinators are able to report frankly to evaluators on the realities of their work, on problems, frustrations and successes. This increases the worth of the evaluation findings.

Figure 1: Evaluation structures of the LdV mobility programme



3. Case study 2: European Community Study Visits programme for those responsible for Vocational Training (CSV)

3.1. General framework

CSV has operated since 1985 under a resolution of the Council of Ministers of 13 July 1983 as an action programme to implement a European Community vocational training policy in the 10 Member States (Belgium, Denmark, Germany, Greece, France, Ireland, Italy, Luxembourg, the Netherlands and the UK).

The accession to the EU of Portugal (1986), Spain (1986), Austria, Finland and Sweden (1995) has significantly expanded the territory of application. The programme was integrated into LdV in 1995 as a supporting measure (Council Decision of 6 December 1994 on LdV I).

That year, Norway and Iceland received permission to participate. LdV II (Council Decision of 26 April 1999) integrated CSV into the mobility policy (II 1.c). Apart from the above-mentioned Member States, in 2003 the future Member States (Bulgaria, Cyprus, the Czech Republic, Estonia, Hungary, Latvia, Lithuania, Malta, Poland, Romania, Slovakia and Slovenia, together with Liechtenstein and Turkey (both under negotiation) will participate, totalling 31 nations.

The European Commission entrusted Cedefop with programme implementation. Cedefop organises CSV in conjunction with national liaison officers (NLOs), who may be supported by national agencies ⁽⁹⁾. Cedefop rules (e.g. participant cost sharing) govern programme budgeting. CSV is a separate Cedefop budget item. Each country usually finances the national support agencies and the many programme organisation expenses (e.g. lecture fees) ⁽¹⁰⁾.

This partial allocation of costs to national budgets explains why countries are interested in evaluation of the segments of CSV organised by their national support agencies.

3.2. Programme objectives

'The prime aim of the programme is to encourage European-level exchanges and discussion among those responsible for vocational training on subjects of common interest.' (Cedefop, 2002, p. 1) ⁽¹¹⁾.

Besides personal familiarisation with a vocational training system in one participating country, other important benefits of the programme are information flow and exchange of experience and views (supranational exchange) between members of the study visit group, who generally represent various stakeholders such as a social partner (e.g. a programme in France with 10 to 15 experts from various countries involved in CSV, excluding the host, France).

The visits are opportunities for participants to profit from the experiences of other vocational training systems, especially from VET innovations. The aim should be to establish international networks and to encourage active involvement in other pertinent activities, such as planning and implementing vocational training policy measures in Member States and other countries taking part in the programme.

The fundamental European policy objective of encouraging reciprocal understanding of vocational training systems, stands over and above these narrower, more focused objectives.

⁽⁹⁾ Each country has its own structure. In many cases implementation is carried out by members of another organisation which is involved in educational programmes and/or initiatives, e.g. LdV agencies, CDG (now InWent), the British Council, etc. Functions and tasks are divided differently between the NLOs and programme organisers in each country in accordance with varying national government structures and policies.

⁽¹⁰⁾ PHARE funds cover grants for participants from future Member States who are members of PHARE programmes. Other countries such as Norway and Iceland must cover participants' bursaries out of their own resources.

⁽¹¹⁾ The catalogue is trilingual (English, French, German). In German this passage is worded: 'Mit diesem Programm sollen in erster Linie Austausch und Diskussion zwischen Verantwortlichen im Berufsbildungsbereich zu Themen von gemeinsamen Interesse auf europäischer Ebene gefördert werden.' And in French: 'L'objectif premier de ce programme est de stimuler l'échange et la réflexion entre responsables de la formation professionnelle sur des sujets d'intérêt commun au niveau européen.'

CSV, a wide-ranging Cedefop initiative aimed directly at the general public, raises the profile of the organisation and indirectly promotes awareness of the EU programmes supporting vocational training in the countries involved.

3.2.1. Programme participants

The programme targets vocational training decision-makers and multipliers. Addressees include government and business representatives, social partners, vocational training institutions and VET researchers. There should be arrangements for the social partners to designate a certain number of participants.

This often rigid 'quota system' is not unanimously applauded. However, the study visit programme is the only programme that addresses the social partners explicitly. For this reason Cedefop and the social partners on the Cedefop Management Board are also keen to ensure participation in this programme by both labour and management, who contribute heavily to vocational training in many CSV countries. EU educational policy requires increased social partner involvement and this is the rationale behind participation 'targets' for these groups.

A working knowledge of at least one other EU language in addition to the mother tongue is desirable. The *lingua franca* will be predetermined separately for each measure but with rare exceptions it will be English. There are 70 measures planned for 2003, four will be in German, one bilingual (German and English), one in Spanish and seven in French. Cedefop offers interpreting support to facilitate communication between visitors and the subject-matter experts in the host country who are not sufficiently versed in the chosen language.

Applications are submitted to the NLOs, who make a preselection and forward the shortlist to Cedefop. The coordinators in Thessaloniki determine the constitution of the multinational group.

Since many months may elapse between application and visit, high cancellation rates are common. Reducing the number of dropouts is a permanent challenge for the CSV central administration at Cedefop, the national support agencies and NLOs.

Cedefop awards participants a grant toward their travel and accommodation expenses ⁽¹²⁾.

3.2.2. Programme procedure and content

Individually themed study visits for groups of 10 to 15 members from the countries involved (see above) with a variety of professional 'horizons', meet in a given country whose residents do not 'participate' in 'home' sessions. Usually a guide accompanies the group as a 'cultural' mediator between the guests and the host country.

The themes are set annually and are valid for two to three years. They are the product of negotiations between the theme selection committee, the NLOs and the Cedefop Management Board in accordance with LdV specifications. The themes are prescribed but NLOs have scope to set national emphases/priorities within the annual theme definitions. Even the 'general' topics (vocational training for young people, adult VET) are usually given a certain focus.

Each CSV country is involved in two kinds of programmes, incoming measures at home and outgoing measures abroad. For instance, in Germany individuals are chosen to join Danes and Spaniards at a measure organised in France. The German programme coordinator organises CSV incoming measures for persons selected in other participating countries. This arrangement ensures that groups contain members from a range of nations.

The national agency ⁽¹³⁾ organises the measure, and the NLO is in charge. In 2000-02, the figures for Germany were:

Table 3: **The CSV: figures for Germany 2000-02 (incoming, outgoing)**

	2000	2001	2002
Outgoing (Germans)	76	71	71
Incoming (other countries)	47	52	72

⁽¹²⁾ Cf. Footnote 11 and Section 3.1 General framework.

⁽¹³⁾ Cf. Footnote 9.

In principle, populous countries organise several measures (e.g. six each in the case of Germany and the UK), while 'smaller' countries often only host one. Member States have organised approximately the same number of programme measures for many years, about 56-59 annually. However, the scale of the programme and consequently the burden on the central organisation at Cedefop are constantly increasing due to the participation of other European countries (Section 1). In 2003 a total of 26 countries will be involved in the planned 70 measures with 7 thematic focuses. If necessary, e.g. for a new field such as 'informal training', they are prepared by having one NLO offer a workshop for all of his/her counterparts.

Normally measures last five days and used to concentrate on the (initial or continuing) vocational training system of the host country. The 1997 meeting in Copenhagen led to the introduction of supplementary three-day ⁽¹⁴⁾ thematic measures. Thematic events have now become standard. The special themes for 2001 and 2002 in Germany were: ICT in SMEs, the quality of vocational training systems and the role of universities in initial/continuing training.

Currently the host country decides on the duration of each measure. Themed events may therefore last four or five days and general sessions on the vocational training system only three days. Variation in the length of measures provides flexibility to adapt to national variants of the vocational training system and to implement European and national priorities.

Specification of themes through discussion and agreement takes place together with a critique of the current programme at annual meetings (Copenhagen 1997, Vienna 1998, Helsinki 1999, Lisbon 2000, Thessaloniki 2001, Berlin 2002) of the NLOs, staff from the coordinating body (Cedefop) and appropriate Commission officials from Brussels, social partners and Cedefop Management Board members.

In 2000 a total of 634 people participated in 63 study trips. Seventeen dealt with vocational training for young people. Thanks to the expansion

of measures as the number of participating countries increases (e.g. accession of newcomers to the EU ⁽¹⁵⁾ and associated countries), participation is growing steadily (from 714 in 2001 to 750 in 2002). This trend is expected to continue.

All participating countries are advised to send their prospective participants to orientation seminars. Their format depends on financial and personnel resources and other circumstances in individual countries. Germany is one country that conducts such seminars. Comparability and transferability of vocational training systems was the theme in March 2003, for example.

3.2.3. Peculiarities, reporting

For several years participants have written short group reports. These are intended to concentrate on analysis and assessment of the theme of the visit, with more general consideration of what was presented and experienced and the organisation of the measure in question. Portrayals of the system investigated are not the purpose, descriptive elements should always be analytical.

3.3. Evaluation of the programme as a whole, individual programmes, particular aspects, etc.

3.3.1. Programme dimension, evaluation stakes and documentation of findings

The CSV programme description, in which it was possible only to outline the developments since its inception in 1985, nevertheless gives an indication of the many parameters requiring evaluation. Programme figures have climbed steadily.

In 2003 the established programme offers almost 800 persons from 31 countries 70 measures in 26 countries encompassing 7 thematic areas. Since 1985 several thousand individuals have benefited. CSV findings are expected to make a major contribution to mastering the challenges currently addressed by the LdV II focus on mobility.

⁽¹⁴⁾ Experts on a certain theme usually only have a limited amount of time and can rarely afford to take a whole week off, including travel to and from the visit.

⁽¹⁵⁾ Future Member States may already participate (linked to LdV participation).

'Programme results' consisting simply of Cedefop, NLO and/or national support agency annual reports can hardly do justice to these figures and to ambitious programme goals such as sharing thoughts, disseminating vocational innovations, international networking and mutual understanding of national VET systems. They merit independent evaluations and supplementary publications documenting programme findings, e.g. participants' reports, workshop presentations, dissemination projects. It would benefit evaluation projects if this aspect, which reflects the complexity of the situation, could be expanded and refined.

3.3.2. Scope of the evaluation and focal points

Evaluation could focus on the entire programme in a given year; on individual types of measures, e.g. a particular theme or the length of the measure, three vs. five days; on specific visitor groups, e.g. social partners, divided into employers and employees, educational administrators or researchers or on a particular partner country. The populous countries involved since the programme began in 1985 would be appropriate because of the greater sample size in both longitudinal and cross-sectional studies. The selection criteria should be based on social science research criteria rather than educational policy considerations such as a 'particularly advanced vocational training system' or 'greater need for reform', in order to evaluate the extent to which the study visits contribute to reflection, further development, etc.

In order to study the process of change, there should also be discussion of whether only one annual programme should be evaluated or whether the preceding periods should be included in the investigation. A distinction should be made between case studies and overall accounts.

Regardless of the scope of an evaluation in terms of scale of programme, the focus of the evaluation must also be defined in programme goals, programme implementation and programme success (dissemination and durability).

For the 70 measures in 2003, each boasting multinational groups of participants, it would also be necessary to explore the following focal points/topics:

- (a) the match between the themes (framework themes and individual thematic focuses) and (CSV, LdV) programme objectives;

- (b) organisation of the overall programme (Cedefop);
- (c) organisation of measures independently conducted by NLOs and/or agencies;
- (d) national selection (free and quota system) and survey of participants (e.g. motivation, match between themes and occupational field, yield and report);
- (e) implementation of general themes in national measures;
- (f) participants' reporting (e.g. analysis of structures observed in the host country, conclusions for their own work) and circulation of reports;
- (g) group dynamics in the individual measures;
- (h) dissemination in the national context and on a European scale,
- (i) further development of a contextualised understanding of confrontational analysis in the host country;
- (j) development of intercultural skills;
- (k) continuing training behaviour following the measure as an outcome of deficit analysis;
- (l) international networking;
- (m) correlation between 'prior educational level' (intercultural understanding, system familiarity, language skills) and success of the programme/measure;
- (n) NLO share in the success of outgoing and incoming measures;
- (o) support given by electronic platforms as a function of programme success;
- (p) role of annual meetings.

It is easy to identify and to appreciate the correlations between three complexes to which most of the above questions can apply:

- (a) national programme implementation (host country);
- (b) national programme participation (participants);
- (c) overall organisation by Cedefop combined with the European dimension.

These could be given concrete form in terms of the focal points/themes cited as illustrations.

Developing the list of themes for evaluation and the consequent design of the investigation requires thorough consultation between the evaluation team and the programme organisers as a basis for decisions on the depth and scope of the evaluation.

3.3.3. Evaluation material, data, procedures and costs

The undisputed merits of evaluation are kept within manageable bounds by the scope and emphases of the evaluation. These should dictate what materials and data form the basis of the evaluation. This selection cannot be made reasonably without sifting evidence such as assessments of organisers' and participants' reports, questioning organisers, interviewing participants (by phone, questionnaires, face to face, etc.) and complex use of qualitative and/or quantitative evaluation tools such as auditing, e.g. path analysis and document analysis.

The available financial backing (evaluation costs) indirectly influences the scope of the evaluation and, consequently, evaluation procedure.

3.3.4. Examples of evaluation

Two examples serve to demonstrate evaluation options for a thoroughly comprehensive and sophisticated programme. One is a national evaluation in a populous country, namely Germany. The second is an overall CSV evaluation. More insight is provided by citing the measures taken by programme organisers to support (continuing) evaluation, i.e. direct integration of evaluation into the programme. Evaluation findings interest a wider circle than the European Commission, Cedefop and the NLOs. Achieving the intended multiplier effect necessitates informing target groups in areas designated for dissemination. This applies particularly to reporting by participants, including results of dissemination projects, etc., and to workshops scheduled to prepare programmes and to establish programme outcomes. Publication procedure will therefore also be included in this case study.

3.3.4.1. National evaluation (Germany 1995-98)

(a) Evaluation goal and operationalised investigation objectives

From 1985 to 1998 a total of 1 341 persons participated in CSV measures conducted by German organisers on behalf of Cedefop and the Federal Ministry of Education and Research (BMBF). This figure shows the significance of CSV within LdV and led programme planners in Germany to reflect on why VET policy-makers' short stays abroad

are vital steps toward the Europeanisation of their field. The planners hoped that CSV evaluation could help to promote transparency and acceptance. Consequently, BMBF commissioned a short-term evaluation of CSV for the 1995-98 programme years, confined to German participants and outsiders attending CSV measures in Germany.

Here is a brief outline of this 1995-98 study. Initial analysis of background material and literature, EU policy papers on the LdV action programme and interviews with German programme managers allowed definition of the goals of the investigation, which were subsequently modified in dialogue with those responsible for the programme (InWent, BMBF). The questions the investigation was to answer were:

- (i) Who represents Germany in the CSV programme and which participants responded to the survey? (What criteria are applied in selecting participants?)
 - (ii) How can we explain the high participant cancellation rate?
 - (iii) What are the motives of German participants in CSV?
 - (iv) How do participants rate CSV programme implementation?
 - (v) Does the CSV programme contribute to establishing contacts, international networking and skilling?
 - (vi) Does the CSV programme perform its function of conveying knowledge about foreign VET systems?
- (b) Evaluation design
- Different methods of investigation may lead to either conflicting or similar results to the same questions. Therefore, an overall approach, analogous to the multitrait-multimethod system, was designed. The model delineates and portrays the ability of various research tools to identify similarities and differences and the thinking behind them. This approach permits internal validation of the interpretation procedure. Statements are internally valid if the use of different research tools produces comparable conclusions.
- (c) Research tools and data sets

Table 4: **Evaluation of the CSV referred to Germany 1995-98: research tools and data sets**

Data sets	Populations
1997 and 1998 questionnaire responses	German participants
Observations (field notes) from 1997 and 1998 preparation and networking seminars	German participants
Observations (field notes) on a 1997 Cedefop workshop in Munich	Multinational group of participants
1995-96 participants' reports on measures in non-German EU countries with German participants	Multinational group with German participants
1995-96 reports by participants visiting Germany	Multinational group excluding German participants

This gave the evaluation three focal points: (i) evaluation of participants' reports and other files/documents; (ii) questioning of German participants; (iii) participatory observation of sample programme implementations.

(i) 1995-96 participants' reports

At the end of their trip, participants were required to draft a short common report to provide an indication of how the study visit had gone, and of its usefulness. Cedefop furnished an outline of desired contents, but this was not binding. Consequently the quality of the reports varied significantly. Some were very short and too general and most were group accounts ⁽¹⁶⁾, the products of a multinational group which offered no scope for an individual's statement. It was not possible to determine whether all participants had always contributed equally to the composition of the report or whether a few 'spokespersons' with linguistic proficiency were the primary authors of the opinions and assessments. Since more than one report was submitted from most countries however, we can presume that the accounts adequately reflected the views of the entire sample. A questionnaire was devised to obtain individual statements, primarily from German participants. The evaluation assessed participants' reports covering 1995 and 1996. A distinction was made between:

- reports by German participants and accounts of groups which included Germans;
- reports by participants in CSV measures in Germany, excluding Germans.

The 1995-96 mandatory reports were subjected to a qualitative content analysis to check the statements drawn from the questionnaires and to answer questions on the implementation and benefits of the CSV programme. The following checklist for assessment of participants' reports was derived from the questions posed by the evaluation study.

(ii) Survey of German participants

A questionnaire was used to measure the expectations, attitudes, impressions, opinions and judgements of former and future German participants. The descriptive and statistical data in the questionnaire responses were analysed, while the characteristics and preconceptions of the participants were treated separately.

The 1997 questionnaires were distributed to former and future CSV participants and covered the following areas:

- critical confrontations and their assumed causes;
- positive interactions and their assumed causes;
- perceptions about participants from different cultures and contacts in the host country;

⁽¹⁶⁾ Participants were asked (by Cedefop) to draw up a group report; this did not exclude individual reports (e.g. by people who might find it easier to express themselves outside the group).

- assessment of networking through the CSV programme;
- positive and negative impressions of the implementation of the measure;
- participants' objectives and major aspects of the measure.

Since, with rare exceptions, only future participants attended the orientation seminar in 1998, the questionnaire was then modified. The 1998 version was divided into five clusters:

- general data on personal characteristics;
- goals related to the CSV programme;
- preparation for the CSV programme;
- networking;
- assessment of and suggestions for improving the implementation of CSV programmes.

(iii) Participatory observation

Observers attended a 1997 networking seminar in Köln-Hürth and a 1998 orientation seminar in the same district of Cologne. Participatory observation of the two events was intended to estimate the significance and quality of these preparatory measures. Participation in the three-day Cedefop workshop in Munich

on funding of training activities served to evaluate the quality and group dynamics of these measures.

Participatory observation is designed to produce the greatest possible rapport between the researcher and the subjects in order to obtain insider perspectives. A semi-standardised procedure is the best that can be achieved. Before seminar and workshop participation, guidelines for observation were drawn up on the basis of questions arising out of the evaluation study. The notes made during participatory observation were analysed qualitatively and presented in the evaluation.

(d) Data analysis and findings were subdivided into:

- (i) Evaluation of organisation and general CSV objectives that covered:
- 25 mandatory reports by the 122 foreign participants in 1995 and 1996 in Germany (a survey of the entire sample),
 - responses to the questionnaire from Germans who attended the two Köln-Hürth seminars,
 - 99 mandatory reports by the German participants taking part in measures in 16 countries, constituting a virtual blanket

Table 5: Checklist for assessment of participants' reports

1	VET aspects
1.1	Can VET lessons be learned from the host country?
1.2	Can the host country learn VET lessons from the report writers?
2	Programme implementation aspects
2.1	Satisfaction with the conduct of the programme (content and organisation)
2.2	Time management
2.3	Opportunity for exchange
2.4	Opportunity to forge contacts
2.5	National coordination
2.6	Funding
2.7	Evaluation of materials furnished to prepare the trip
2.8	Establishment of contacts, networking
2.9	Multiplier effects
2.10	Suggested improvements

survey, with the exception of Germany ⁽¹⁷⁾,

- findings of the participatory observation of the two Köln-Hürth seminars, (e) the Munich workshop.

The investigation goals, the guidelines for thematic analysis referred to above and the questionnaire clusters were translated into the following focal points:

- content and composition of the programme;
- relevance to the theme of the study visit;
- accommodation of individual wishes;
- style of lecture presentation;
- allowances for participants' linguistic ability;
- written information material;
- time management;
- opportunity for exchange and reflection;
- opportunity to forge contacts;
- national coordination;
- hospitality in the host country;
- funding;
- materials furnished to prepare the trip;
- networking and contacts made, initiatives for future cooperation;
- multiplier effects: further sharing of impressions and insights.

The findings from this evaluation, covering 1995-98, are summarised below. Learning about another VET system on the ground was the main incentive. Therefore, in addition to the tours and presentations, participants wanted more time for mutual comparison of vocational training systems, for sharing experiences and for discussing new impressions. VET systems were portrayed by lectures, tours and discussions with those involved in vocational training. Constant alternation between these media was greatly appreciated. The predominance of lectures was criticised.

Opportunities to forge contacts and to reflect on experiences alone or as a group determined whether a visit was a success or a failure. Participants could have made

contact before the trips if a list of participants' addresses had been circulated in advance. Devoting insufficient time to introductions at the beginning of the tour made it harder to make contact and identify common interests. Respondents indicated that chances for interaction were on the whole inadequate.

Information to help people plan their trips was often insufficient and frequently arrived too late. Those polled welcomed the preparatory meetings in their home country. Many respondents stressed the importance of preparing by reading and attending seminars. The preparatory and networking seminars were highly rated because of the interaction between future and former participants. Follow-up seminars offering opportunities for post-mortems were rated as highly as preparatory measures.

Daily conferences were proposed to boost networking. Since many participants had not previously been on an exchange, they lacked experience in making international contacts and requested help in this area. Group dynamics within tour groups were consistently considered positive. They had led to the planning of more joint activities in future.

The lack of cultural activities was frequently criticised. If offered, they were valued highly, particularly because they provided an opportunity for informal contact.

This general evaluation also investigated German participants. Their average age was 44. More than half were men. One quarter had already been abroad on other exchange programmes. The participants were highly qualified. Most were university graduates and held high-ranking posts. Only one in four had been through a selection process.

(ii) VET evaluation of the CSV programme.

Qualitative hermeneutic text analysis was applied to all reports of 1995 and 1996 EU study programmes in which German participants had been involved or which had taken place in Germany. The vast

⁽¹⁷⁾ Cf. first bullet

majority of CSV measures were, therefore, evaluated. The summary refers to the findings of the expert evaluation of programmes carried out at that time in 17 countries: Austria, Belgium, Denmark, Finland, France, Germany, Greece, Iceland, Ireland, Italy, Luxembourg, the Netherlands, Norway, Portugal, Spain, Sweden and the UK.

Evaluation of professional VET aspects focused on whether the CSV programmes fulfilled the aim of conveying knowledge of other VET systems.

It would have been an unacceptable simplification to base an analysis merely on this aspect, because the complexity of the matter to be evaluated would not then have been taken into consideration. Although the participants' statements initially appeared to cover only vocational training in the host country, they were actually referring indirectly to vocational training in their homeland. Ultimately there was always an implicit comparison of VET systems and their components, since value judgements require a benchmark. Elementary issues of comparative vocational training research were thus raised. These focused on three major areas:

- becoming acquainted with VET structures in other countries and examining their characteristics in order to understand these and to explain differences between individual systems (idiographic function);
- overcoming the melioristic approach to comparison as a quest for the best system and abandoning ethnocentric stances on VET issues;
- recognising common problems and trends (evolutionist function) in various countries, identifying common regulatory provisions (nomothetic approach) and searching for universals (quasi-experimental function) between systems.

Differing national environments markedly shape the educational and vocational training systems of Member States. Despite this diversity, there are a number of common problems and trends.

According to CSV programme the first two objectives are particularly relevant to evaluation. The third requires a more profound knowledge of systems and is the subject of comparative educational research. The investigation also checked whether programmes addressed problem areas and whether participants commented on development trends.

The obligatory reports were condensed, brief and rather general. This ruled out content analyses of the overall VET system. However, it was possible to find responses which:

- commented on the new experiences as a contribution to the debate on reform in the respondent's own country (idiographic function);
- were critical of the host country or adopted an attitude of knowing more about vocational training than was found there (ethnocentric stance);
- described problem areas and drew conclusions on trends (evolutionist function).

These statements are indicators of knowledge of the host country VET system and of reflection on the differences between the 'home' system and 'their' system, or on sub-themes which were highlighted in the implementation of the programme. Since supplementary interviews with previous participants and questionnaire surveys on new professional insights, akin to the multitrait-multimethod approach, were beyond the scope of this brief evaluation and would have exceeded the budget, statements from participants' reports were the chief criterion for evaluation of 'system understanding'. Document analysis, therefore, focused on whether the author of the report was able to learn from the host country (idiographic function); or 'knew better' than the host country (ethnocentric stance).

There was also examination of whether the reports had a bearing on problem areas or identified trends (evolutionist approach). This document analysis was not very fruitful because most of the report writers concentrated fully on the

programme as they had experienced it. The group reports were 'balanced' compositions.

The findings of the VET evaluation were ambivalent. Cedefop guidelines for the structure and content of the reports evaluated were not strict enough. The reports ranged from comprehensive treatises with very detailed contents and in-context critical remarks to short findings limited to a few sentences that provided little scope for assessment. These statements were frequently too general, unfounded and, therefore, not conducive to coherent overall analysis.

Many of the presumptuous suggestions, which often assumed that there was a 'better' VET system than the structure studied in the host country (usually the author's own establishment was implied), proved to be based on misunderstandings or insufficient information. Unfortunately, only slightly over 50 % of participants exhibited full, unbiased understanding and approval of the 'foreign' VET system. Better recognition in national programmes of the different circumstances of participating countries through 'enlightenment' about the 'foreign' systems discovered and combating of prejudices may significantly help to improve transparency and programme outcomes.

- (e) Recommendations and model programme
Evaluation findings were translated into recommendations with the following focuses:
 - (i) networking,
 - (ii) linguistic competence,
 - (iii) methodological and didactic planning and implementation of study visits,
 - (iv) a complex of requirements for successful exchanges,
 - (v) professional assessment of the implementation of programmes (was knowledge of VET systems transferred and did participants gain professional inspiration?),
 - (vi) participants' reports,
 - (vii) areas of conflict between recommendations and responses to the often-heard charges that study visits were nothing but 'tourism'.

They were also shaped into a model programme (preparation, implementation, follow-up, outcomes and requirements), in the full knowledge that national implementation is based on the subsidiarity principle.

- (f) Dissemination of evaluation findings
The recommendations and the model programme constituted the basis for a workshop with the organising agency and the NLO. Another result was the issue of brochures on *Financing of vocational training in Europe* and *Vocational training in European dialogue*.

In its capacity as central coordinating agency, Cedefop supplemented studies of national organisers with its own evaluation of findings. Numerous measures were planned and some were implemented. Several were included in the following description of evaluation measures.

3.3.4.2. Overall evaluation of CSV

In 2000, CSV was appraised as part of a Cedefop evaluation. The main purpose of the evaluation was to assess the VET aspect of CSV in relation to Cedefop tasks, its role in LdV operations and in the promotion of EU vocational training policy. Since this evaluation did not directly address CSV objectives, it required different data sampling. Investigators collected data by phoning and interviewing policy-makers and organisers (Cedefop's CSV department staff, representatives of the European Commission and of international interest groups and associates of the Cedefop Management Board) and by sending questionnaires to stakeholders. The programme goals and the impact on participants were not the focus of the Cedefop study. Instead it concentrated on programme organisation from the standpoint of Cedefop and NLO organisers, external impact and the dissemination of programme findings in Cedefop activities.

Although the study testified to the great commitment and organisational talent of those involved in implementing the programme, this evaluation, which can more accurately be described as a consultation about organisation, neglected the inherent purpose of evaluation, which should focus on programme goals, participants and subsequent impact. Instrumentalisation in the interest of EU vocational training policy and Cedefop operations in fact negated the huge

potential of CSV measures to encourage participants and persuade them to adopt programme goals. The bottom-up effect, which would have fostered the long-term impact, was sacrificed for a top-down effect. Such a summative evaluation addressed neither programme aims and participants nor the organisation of the programme in achieving the programme's purpose. In the final analysis, it did not do justice to Cedefop organisers, NLOs or participants. This claim can easily be proved because measures initiated by Cedefop to improve the programme are particularly designed to reinforce bottom-up effects.

3.3.4.3. Measures to support (continuing) evaluation

Cedefop CSV coordinators and numerous NLOs obviously consider formative evaluation to be essential to further CSV development. Since 2002 an appropriate action has been taken to stress preparation for measures, reporting, networking and dissemination. Participants receive a questionnaire, which focuses not only on organisational issues but also on networking and dealing with innovations in vocational training. NLOs also implement participant monitoring by means of questionnaires, all evaluated centrally by Cedefop. This toolkit, devised by an independent team of social scientists, forms the basis for formative (self-) evaluation. Implementation has been supported by numerous task forces made up of NLOs, national agency delegates, members

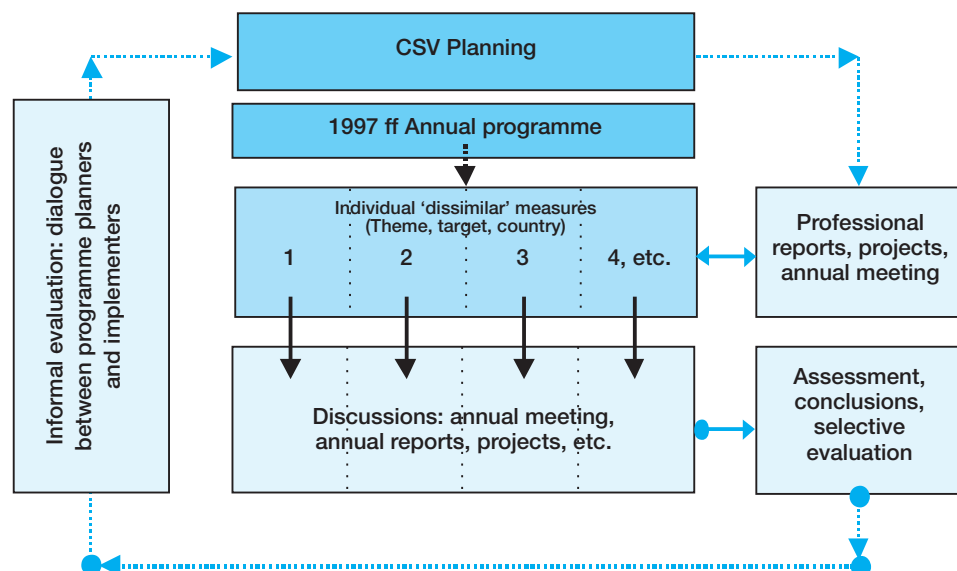
of the Cedefop CSV department and CSV stakeholders such as social partner representatives on the Cedefop Management Board.

This 'self-evaluation' is meant to be augmented by an external evaluation. Apart from a meta-evaluation of the self-evaluations, a follow-up survey of all participants is intended to provide a gauge of the lasting impact of measures. A pilot study was conducted in 2002.

3.4. Summary

The CSV example lends itself to depicting the evolution of the culture of evaluation. Up to the mid 1990s evaluations resembled annual reports with numerous tables of statistics which said little about programme quality. In Germany, the NLOs (representatives of the Ministry of Education and Research) view that the potential of CSV should be enhanced to meet programme goals, led to an evaluation of data, amounting virtually to a complete survey, provided by the NLO and the implementing organisation. The findings of this formative evaluation, combining qualitative and quantitative methods, were recorded in an expert report. Since programme revisions without input from the scientific community would obviously be seen as top-down decisions, a booklet containing evaluation findings was sent out to specialists.

Figure 2: 2001 CSV evaluation structure before introduction of new evaluation practices in 2002-03



Since then, the organisers involved have been engaged in lively debate with the European Commission, the NLOs, the social partners and members of the expert community, to name just some of the most important participants. The findings manifest themselves in a new organisational, reporting and dissemination structure, which supports continuing formative evaluation but also permits self-evaluation. Participants and stakeholders, such as the social partners, are actively

involved in this approach to integrate formative evaluation into CSV programme structures.

CSV can thus demonstrate how programme structures and evaluations can be developed and aligned through what is ultimately a political process. The JCSEE and DeGEval criteria formed the terms of reference for all four cases addressed in this study ⁽¹⁸⁾. In this context CSV is a relevant example of the programme development process.

⁽¹⁸⁾ Cf. Section 1.3.2.

4. Case study 3: the German-French Youth Foundation (DFJW)

4.1. General framework

The German-French Youth Foundation (DFJW) was founded on 5 July 1963 as an organ of Franco-German cooperation, following an agreement between the foreign ministers of the two countries. DFJW's self-appointed task is to promote youth exchange between youth organisations, sports clubs, language centres, vocational training institutes, professional associations, trade unions, schools, universities, communities and twinning associations. DFJW's main role is that of mediator. It offers support and advice on financial, educational and linguistic aspects of exchanges. The foundation also informs and guides providers on the content of meetings and follow-up analysis. It is financed jointly by the German and French governments and public and private organisations. Government assistance for 2001 totalled approximately EUR 20 million. Each year about 200 000 young people participate in exchanges and around 7 000 programmes are organised. Since 1963 more than 250 000 visits have taken place, involving over 7 000 000 young people.

In addition to the binational programmes, DFJW supports increasing interaction with other countries (in Central and Eastern Europe, South Eastern Europe and the Balkans, the Mediterranean, etc.). Currently 350 trilateral visits are arranged each year.

4.2. Programme objectives

DFJW supports many intercultural exchanges in several different thematic areas. 'Careers and training' is a particularly important example.

Increasing internationalisation is reflected in labour-market demand for long-term 'skills for Europe'. Young employees are expected to have greater occupational mobility, familiarity with other countries and cultures and fluency in foreign languages. Occupational mobility depends largely on key skills. These include linguistic, intercultural and methodological proficiencies. It is not only a matter of young people acquiring these skills: exchange programme supervisors must also improve their knowledge in these areas. Stays abroad are becoming increasingly important components of the vocational training of young people and stepping stones for their entry to the labour market (currently hindered by high unemployment levels).

DFJW addresses these developments in its work on Franco-German partnerships, cooperating with various training institutions to focus activities on promoting these crucial career skills. The programmes are directed towards both individuals and groups. Special projects have also been set up for young unemployed people and the socially disadvantaged. Examples of initial vocational training projects are partnerships between chambers of craft trades. These allow young trainees to view their trade from the perspective of another country with a different culture and education system⁽¹⁹⁾. The aim is to give them additional perspectives on their profession and the opportunity to experience other cultures at first hand.

DFJW also collaborates with the Danish PIU (*Praktik i Udlandet*) Centre on the vocational training initiative TransNet. TransNet is a European network of institutions which organise and support work placements abroad. The placements are temporary job-related activities which

⁽¹⁹⁾ Three intercultural projects for young artisans:

- most chambers of craft trades in both countries are involved in Franco-German partnerships and organise exchanges between young artisans with the assistance of DFJW. Visits normally last from 10 to 14 days. German trainees become acquainted with working methods in French companies and with the people and the culture of their partner region;
- DFJW assists young artisans to participate in one-month 'short placements' in France during their training;
- trainees who wish to work in a French craft trade firm for at least six months after their training contract has finished receive twofold assistance from DFJW. This pays both for a language course before the period abroad and for travel. Wages are set according to the host firm's policy.

young people engage in during initial vocational training, in the first few years in work, while studying or taking further vocational training courses or as training when unemployed. The objectives are to improve communication and cooperation between local, regional, national and international institutions which organise or support work placements abroad and to analyse and help to overcome mobility obstacles to placements abroad for school pupils, young employees, young unemployed people who have completed initial training, students and recent unemployed higher education graduates.

One of DFJW's aims for the programmes involving third countries is to pass on to third countries the experience of reconciliation acquired through Franco-German youth exchanges.

4.3. Evaluation of the overall programme, individual measures, particular aspects, etc.

Examination of DFJW activities is particularly relevant to our present purpose since it offers a profound insight into how international exchange programmes are evaluated. The DFJW programme itself has generated a considerable body of literature, which has gone beyond its immediate aims and which significantly contributes not only to research into intercultural mobility and exchanges but also to the evaluation debate. The DFJW's long history of evaluating its programmes has produced many approaches which differ in both their procedures and their focus. Numerous statistics, annual activity reports and inventories of achievement have illustrated the foundation's successes since its creation 40 years ago. The activity reports record the number of participants sponsored annually and their ages, nationalities, educational levels, etc. The report produced in the year of the programme's 30th anniversary (DFJW, 1993,

p. 61 et seq.) states that, following a drastic slump in the 1970s, participant numbers had been climbing continually since 1980 and had again reached earlier levels. The 10 % increase in the number of participants recorded in 2001 is attributed to the programme's extension to a wider clientele. Statistics such as these are updated and published annually.

The foundation enhances its reports by regularly commissioning quantitative investigations ⁽²⁰⁾. These usually take the form of surveys. Past studies have addressed topics such as the changing opinions of German and French youngsters and young people's awareness and appreciation of DFJW.

DFJW's guidelines stipulate that all its programmes must be evaluated ⁽²¹⁾. The requirement is fulfilled by group reports on the quality of activities, a report from the management team and other assessment devices (surveys, tests, etc.).

These reports, which are written on nearly all international exchange programmes, are particularly useful for the evaluation of individual programmes. Instruments of this nature are always influenced by the subjective interpretations and perceptions of the authors, who are ultimately assessing something which they themselves are party to. However, this reporting procedure can be improved and enhanced so that reports are standardised and more objective through the introduction of reporting criteria or participatory observation by persons outside the programme. Both approaches have been adopted in the evaluation procedure.

Over the years, numerous assessment reports have been written about supported providers' exchange programmes and experimental and research programmes. The research-oriented further training programmes conducted in cooperation with associations and research institutes are usually designed as cycles of three or four visits over three years (with a 6 to 10-day visit each year). This is applied research with and into binational and multinational groups of approximately 30 people. Critics complain that the

⁽²⁰⁾ One example is a representative survey of 804 French youngsters and 866 young Germans carried out in 1985 by EMNID and a French partner.

⁽²¹⁾ 'Each programme shall be jointly assessed. This can be achieved via a group report, a report from the management team or by other forms of assessment (surveys, tests, etc.)' (extract from the DFJW guidelines, 2002b, p. 3).

findings of these experimental programmes can rarely be transferred to real-life situations. However, they serve to facilitate more precise description of actual exchange programme processes and can demonstrate their specific characteristics and potential.

Findings from specialist conferences and congresses staged fully or partially by DFJW are also available, in addition to the quantitative evaluations and reports on individual programmes, participatory observations and research-oriented programmes ⁽²²⁾. A series of texts and books and many articles which have been published on the practical experiences of exchange programmes also consider evaluation issues. In general, DFJW follows a clear goal of formative evaluation. This corresponds to one of the tenets in the preamble to its guidelines:

'These guidelines shall be regarded as pointers to practical action; they must accommodate the interests of young people and be adapted to reflect changes in these interests on the basis of continuous evaluation of programme results' (DFJW, 1994, p. 3).

This accords with the demand for continuous dialogue between providers and the foundation about the evaluation results for individual projects, in the aim of continually developing and improving programmes.

These efforts and the work in recent years of DFJW and external bodies on quality standards and evaluation in international youth services have produced a relatively broad consensus on several basic issues. For example, it is apparent that there is no automatic mechanism for meeting people and getting to know them, indicating that the potential of intercultural interaction has not been fully exploited. The increased number of trained mediators and moderators should help overcome this as they are familiar with the cultures involved, have adequate language skills, know about bicultural and multicultural dynamics and have ample practical experience with these group processes. Employing these support staff incurs additional cost, but the benefits of an improved exchange situation far outweigh this drawback. A further programme improvement would be to develop and publish glossaries of

specialist vocabulary in order to facilitate work-related communication and to make greater use of project work, interdisciplinary procedures and interoccupational initiatives based on experiences generated by craft trade work. The increased role that young people are playing in designing programmes and making decisions is a further consequence of assessing previous programmes.

DFJW also has a stake in quality assurance through the TransNet project mentioned above. This has set itself the goal of providing details about international placement exchanges and sharing experiences with all interested specialists via the Internet. The foundation also plans to define quality criteria for international placement exchange schemes and thus contribute to enhancing the quality of participants' stays abroad.

However, DFJW is not immune to the problems of putting theoretically determined principles into practice. For example, it has not been possible to make full use of the knowledge gained by DFJW and other bodies on improving the quality of visits because of the additional burdens on the responsible institutions. It would therefore seem imperative that DFJW implement quality policies by working closely and sharing tasks with its partners. This would involve the foundation taking on a more active service role, providing advice for its partners.

4.3.1. Internal evaluation analysis

A striking feature of DFJW activities is the benefit derived from scientific analysis of the programme, its content and methods. This was clearly shown at a study conference on the theme of *Intercultural relations and intercultural learning* in 1990 and in the 2001 publication *Europakompetenz lernen* by Demorgon et al.

4.3.1.1. *Intercultural relations and intercultural learning study conference*

The conference was jointly organised by the Social science study group on international problems (SSIP), the Exchange research working group, the *École normale d'instituteurs* and the DFJW. It focused on the need for intercultural

⁽²²⁾ A good example is a publication on a study conference in 1990 with French and German researchers on the theme of 'Intercultural relations and intercultural learning' (Thomas and Abdallah-Preteuille, 1995).

exchange and the form that this should take. From experience of the DFJW programme, the conference deduced that the mere availability of opportunities for intercultural meetings was not sufficient to ensure the intended positive and lasting outcomes. Intercultural encounters can have various negative effects such as the denial or trivialisation of cultural differences. This is often because organised intercultural visits can easily acquire the character of a halfway house, starkly isolated from the real world of cultural clashes. If a visit is not professionally organised, misunderstandings can lead to disapproval, dislike or even premature departures. On the other hand, modern, romantic 'multi-culti' aspirations are also doomed to failure if they cannot be properly channelled. Gaining a better picture of the problem and finding the right approach to designing successful intercultural visits demands more precise and thorough analytic approaches which take into account all cognitive, affective, behavioural and personal factors. Only programmes of this complexity will make best use of intercultural exchanges.

The conference also revealed that the German and French programme objectives differ fundamentally. The Germans are interested in the value of exchanges as an opportunity for Germans to gain skills in international interaction. The French focus more on questions of their own cultural identity. They prefer more hermeneutic, symbolic-interactive and ethnomethodological research approaches, whereas the Germans follow positivist deductive traditions, oriented towards action theory and strongly influenced by Anglo-American emphases. This has consequences for the design and implementation of evaluation, particularly since those involved are often not aware of these differences. Conflicts over objectives and lack of mutual understanding are dangers likely to affect quality assessments of international programmes, such as those of the EU, particularly given the recent trend of greater interdisciplinarity.

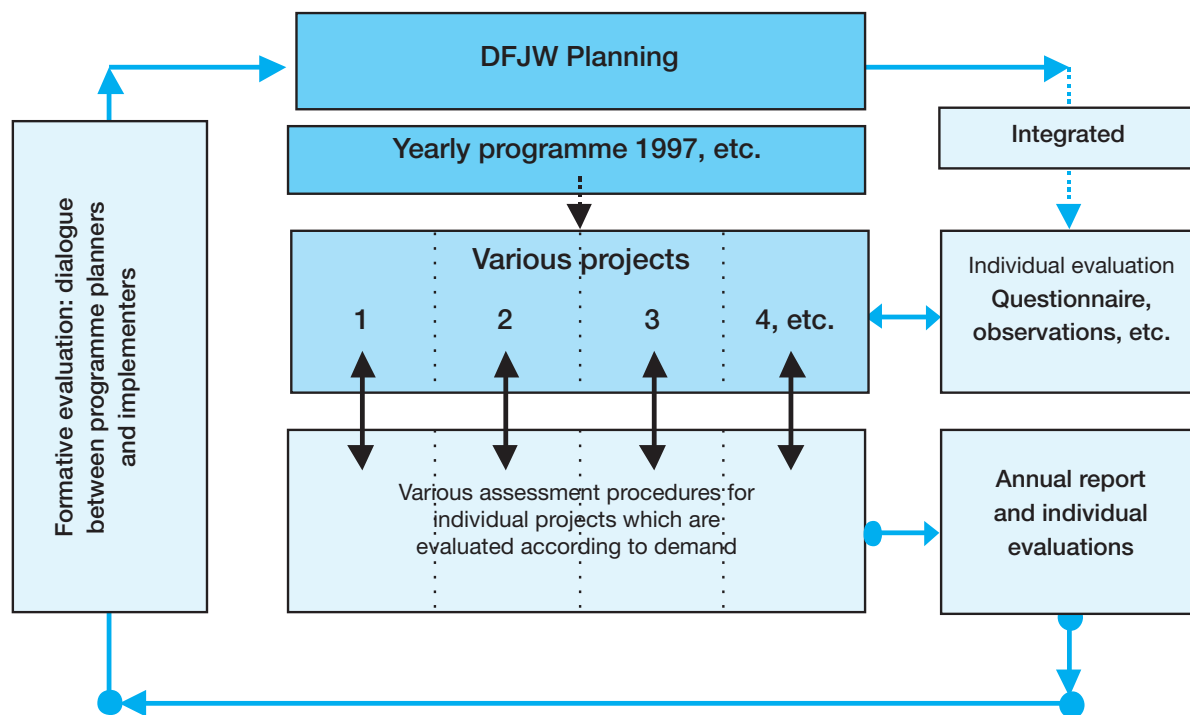
4.3.1.2. *Acquiring skills for Europe*

DFJW's evaluation approach has much in common with the theory propounded by Demorgon et al. who have been involved with DFJW for decades. They advocate a comprehensive approach to evaluation that is closely linked to the topic. Evaluations of intercultural visits

cannot be based on other evaluation models but must derive from the matter to be evaluated. The desire to evaluate intercultural work is, however, paradoxical, in the sense that evaluation presumes the existence of generally agreed standards. This problem is aggravated by the large number of evaluations. In fact, an evaluation of evaluations is necessary to establish which methods can actually provide useful and, more particularly, reliable information. However, in order to ensure reliability and a certain standard of quality, a definition of 'objective' criteria is essential. However, this is particularly difficult in international youth work, since cultural and local differences manifest themselves. It is precisely these differences that give activities their intercultural character. This affects both the search for reliable standards and assessing educational methods, including evaluation methods, which are not culturally neutral but which reflect national traditions, value judgements, etc. The foundation, however, has not reacted to this problem by abandoning its clearly defined demand for quality and resigning itself to inconsistency. Instead, it has made it a goal to establish a culture of communication on assessment criteria and debate on quality and approaches to quality criteria. The chosen means to achieve this is 'self-assessment', which involves programme managers from the different cultures working together to develop categories and criteria for interpretation which they then apply to the arrangements for visits. The advantage of self-assessment is that it can be conducted while a programme is running, permitting findings to be directly incorporated and necessary changes to be immediately introduced. In practice, international exchange visits at all levels represent a continuous process of reciprocal testing, probing and adapting, in a process of practical research and evaluation. External evaluations on the other hand, do not so much serve the interests of the organisation itself as provide a basis for comparisons with other programmes.

The evaluation structure depicted below illustrates how the different aspects of the evaluation process interact and how they are expressed in plans for future programmes. It is clear that formative evaluation is the central idea and that the whole structure is designed to facilitate continual refining of overall planning and the design of individual measures.

Figure 3: DFJW evaluation structures



4.3.2. Future planning

At its board of management meeting on 22 June 1999, DFJW presented a guideline report for the following three years. Alongside an analysis of the current situation, the report considered future prospects. In relation to evaluation, the foundation proposed to build on self-assessment by drawing up notes for the general public on how to design and carry out this type of assessment. It also intended to use the Internet more intensively as a service tool and to publish information on quality issues, particularly examples of programmes as a stimulus for organisers. It also planned to do more to evaluate the increasingly significant trinational programmes, particularly with countries in Central and Eastern Europe and to go on to prepare guidelines describing typical problems and ways to address them.

4.4. Summary

Because of its long history, DFJW has had much experience with many kinds of evaluation methods. However, the different types of evaluation are linked to varying information requirements.

Although it is not entirely possible to prove the effectiveness and success of the programme with a list of annual figures, surveys and performance statistics based on quantitative data, these do help to describe the foundation's activities so that they can be discussed at a general level. At the same time, while subjective individual reports may not give a useful overview of programmes, or measure or prove anything, but they make an important contribution to formative evaluation, since personal experiences can provide helpful hints for possible improvements in the planning, implementation and content design of visit schemes.

Certain aspects of DFJW's evaluation procedure need to be discussed in the light of the evaluation requirements set out at the beginning of this chapter. DFJW is taking the opportunity to engage in scientific debate about issues arising from international visits. This allows it to seek a theoretical basis for its activities and gain an outsider's perspective on its processes. Other positive features include the research-oriented projects, which directly target certain aspects under controlled conditions. In this way DFJW can continue to look for appropriate standards of

quality, which are crucial for impact research. However, there is still scope and a need for further discussion. Standardising the design and practice of individual assessments is certainly a step in the right direction towards increased control over what features individual reports should contain.

As the foundation itself has stressed, it needs to make better use of existing evaluation models

and published research on the topic. It should also improve its organisation and cooperate more closely with its partners. Using binding standards for individual reports, data collection, etc. would make evaluation more efficient and useful. The foundation should also satisfy the requirement for meta-evaluation in order to compare various findings and draw useful conclusions for future programmes.

5. Case study 4: International Exchange Programme for VET Specialists (IFKA)

5.1. General framework

In 1969 the Federal Republic of Germany and Japan agreed on an exchange programme for specialists. In 1978 this became IFKA at the instigation of the German Education Minister, Helmut Rohde. It is now the longest-running vocational training exchange programme of the Federal Ministry of Education and Research (BMBF) and grew out of the political principle of promoting equality between academic and vocational education. Implementation of this principle concentrated on in-company training personnel, who were a primary concern of politicians and social partners, particularly after the VET Act with its regulations on qualifications for trainers was passed in 1969. Most of these trainers were skilled workers or master craftsmen and these are still, after 30 years, the key figures. Consequently, programme objectives are still specifically tailored to the personal development of in-company trainers. With these guidelines, the programme implementation explored the host country's educational and vocational training system in relation to a specific project against the background of cultural, historical, societal, political and economic conditions.

BMBF's proposed new exchange scheme, 'More international mobility in Vocational Training', which has been developing since the late 1990s, maintains this programme objective but enhances it with new emphases. These reflect the increased internationalisation and globalisation which has occurred in the 20-plus years since IFKA was founded. They build upon the expansion in training for vocational training personnel (outside vocational schools) which has taken place since 1969. Projects which had previously focused on the personal development of individuals have been modified for practical purposes to facilitate the transfer of vocational training innovations from

exchange partner countries to Germany. The aims of this process include:

- (a) giving a more international focus to vocational training in Germany;
- (b) ensuring that acceptance of formal and informal qualifications obtained abroad is not hampered by bureaucracy;
- (c) promoting the comparability and transparency of qualifications by establishing equivalences.

To respond to these new extended requirements, the programme has, since 1999, adopted special focuses, tailored to each target country. Some examples are: making vocational training more flexible; IT skills systems; multimedia learning; modularising vocational training; regionalising vocational training; training courses providing double qualifications (equality of academic education and vocational training); training in the service sector; transition processes in Central and Eastern European States; and vocational training system innovations.

When IFKA was founded in 1978, the former Federal Ministry of Education and Science (BMBW) commissioned the Carl Duisberg Society ⁽²³⁾ to conduct the exchange programme for VET specialists. Specialists are assigned to other countries and received as guests according to guidelines laid down on 9 February 1984 by BMBW's successor, the Federal Ministry of Education and Research (BMBF). BMBF stipulates which countries can be involved in the programme and decides on the German applicants proposed by the CDG/InWent expert committee, who apply through a public recruitment process. The foreign partners choose their participants according to their own guidelines.

In general, IFKA is based on reciprocity. An exchange takes place every year with each partner country. 'Outgoing projects' refer to visits involving participants from Germany going

⁽²³⁾ Since November 2002 InWent – Internationale Weiterbildung und Entwicklung gemeinnützige GmbH is an organisation for international human resources development, advanced training and dialogue. It was established in November 2002 through a merger of Carl Duisberg Gesellschaft e.V. and the Deutsche Stiftung für Internationale Entwicklung (German Foundation for International Development).

abroad, 'incoming projects' to visits by foreign groups to Germany. By the end of 2001 over 6 000 German and foreign specialists had taken part in more than 450 information and training visits (15 outgoing projects per year) in 40 countries. BMBF and the foreign counterparts of CDG/InWent directly responsible for the exchanges had invested almost EUR 35 million in the scheme. This does not include the expenses of the companies, associations, etc., involved in exchanges in connection, for example, with hosting company visits, giving lectures and presentations. If these were included, the total amount would probably exceed EUR 50 million.

5.2. Programme goals

5.2.1. Objectives

In 2002, the professional and personal objectives were as follows: 'Education, economic policy, professional and personal matters with the following focuses: continuing vocational training; familiarisation with education and vocational training systems and professional practice in other countries and their economic and political contexts; support for multipliers who pass on the knowledge they acquire abroad to their own companies and others; a contribution to making vocational training systems more flexible and improving intercultural understanding' (extract from BMBF guidelines).

Further training abroad should ensure, among other things, that:

- (a) companies and other institutions sending workers abroad on a IFKA visit profit from the international exchange and the establishment of national and international networks. This applies particularly to small and medium-sized enterprises;
- (b) participants gain additional professional and general skills, including international and intercultural competences;
- (c) experience of a foreign vocational training system enables participants to stimulate innovation within the German vocational training system;

- (d) international mobility and cooperation in education and trade is promoted, strengthening Germany's position as a provider of training in the training market.

5.2.2. Participants

Participants come from backgrounds which fall in the broadest sense into the domain covered by the VET Act (BBiG). These include: vocational trainers, training masters, managers and advisors; apprentice counsellors; in-company continuing training personnel and trainers from approved providers; personnel managers and staff development officers. Prerequisites for participation are several years' experience in vocational training and involvement in inter-company vocational training activities such as training or examination committees.

IFKA is a group scheme with approximately 15 participants per project. They are prepared for their stay abroad by a briefing seminar. Their visit lasts about two weeks during which they learn about the vocational training system of the host country in a series of lectures, seminars and practical demonstrations in companies, vocational schools and other vocational training institutions, etc. The briefing and debriefing seminars are held in Germany.

5.2.3. Use of electronic media for preparation, implementation, dissemination and lasting effect

To prepare participants for their stays abroad electronic media are used increasingly to facilitate communication among them, to create lasting networks and to consolidate results (e.g. innovation transfer and reports on special topics from the host country). CDG/InWent has developed an Internet platform for these purposes.

5.2.4. Partner countries

IFKA exchanges have been organised with the following States: Argentina, Australia, Austria, Belgium, Brazil, Bulgaria, Canada, Chile, the Czech Republic, China, Denmark, Finland, France, the former GDR ⁽²⁴⁾, Greece, Hungary, Indonesia, Ireland, Israel ⁽²⁵⁾, Italy, Japan, the Republic of Korea, Luxembourg, Malaysia, the Netherlands, Norway, Poland, Portugal, the

⁽²⁴⁾ Independent programmes, which adhered to the same principles as IFKA until 1998.

⁽²⁵⁾ Idem.

Russian Federation, the Slovak Republic, Slovenia, South Africa, Spain, Sweden, Switzerland, Taiwan, Thailand, Turkey, the UK (England and Wales/Scotland), the US and Vietnam.

5.3. Evaluation of the programme as a whole, individual programmes, particular aspects, etc.

5.3.1. Principles and focuses

In order to achieve a comprehensive assessment of IFKA, the basic questions asked in the evaluation focus on two issues affecting the impact of IFKA (the transfer of results) on:

- (a) People
 - (i) on programme participants (cause-and-effect model);
 - (ii) on other people or institutions connected with IFKA in some way (dissemination of results).
- (b) Systems
 - (i) on the vocational training structure of the enterprises where IFKA participants work;
 - (ii) on vocational training policies outside those enterprises;
- (c) Programme organisation and its relationship to and influence on 'people' and 'systems'.
Until the mid 1990s evaluation focused on the classic programme objectives related to personal development. Since then, evaluation has increasingly concentrated on IFKA's new 'professional' programme objectives oriented toward vocational training policy.
A distinction has to be drawn between summative evaluation as a global, summary assessment after a year's programme of outgoing projects and formative evaluation

primarily targeted at improving project design. However, IFKA's summative evaluations of annual programmes (e.g. 1999, 2000, 2001) have had a strong formative effect because proposals from previous years' assessments have often been incorporated into exchange programmes in subsequent years. This approach of asking for proposals for structural change as a major element of evaluation was not merely the evaluators' idea, but was also specified by the client. CDG/InWent provides preparation for evaluation by arranging joint workshops between organisers and evaluators and offering advisory discussions.

For practical reasons, there has been no planned monitoring of individual projects as part of an annual programme. A viable scheme would involve evaluators supervising up to 20 measures for 14 days each in many host countries (Section 5.2.4). This is not practicable purely on grounds of cost. In order to demonstrate to evaluators how a programme is implemented in practice, they should consider some 'sample' monitoring of a group, in the role of tutor for example. Several evaluators have taken advantage of this opportunity.

5.3.2. Chronology of the German Institute for International Educational Research (DIPF) evaluation of IFKA

CDG/InWent and DIPF began working together in 1982. They agreed to conduct an evaluation of IFKA participants' reports from 1979 and designed the study together. In 1984 a reader was published on vocational training abroad from the trainers' viewpoint. This focused on the four emphases of subsequent evaluations mentioned earlier:

- (a) programme objectives (proposed and achieved);

Table 6: 1979-83: documentation of findings and formative evaluation

Time period/ countries	Focuses	Type of assessment	Documented in:
Until 1983: Outgoing and incoming, all countries	Organisational framework; career benefit; personal benefit; professional benefit.	Qualitative document analysis with question matrix	Lauterbach, 1984

- (b) programme implementation;
- (c) programme yield (added value);
- (d) proposals for continuing programme improvement.

Proposals for programme improvement resulting from the qualitative evaluation of the participants' reports (at that time freely formulated) initially concentrated on the preparatory stage of programme implementation. They focused on:

- (a) professional familiarisation with the vocational training system in the host country in the context of the general background (the education system, politics, culture, the economy, etc.);
- (b) the organisation of the programme and the measure (especially guidance for participants from support staff, the match between programme objectives and the actual programmes of the individual measures);
- (c) the structure of the programme (exploitation of opportunities to learn about the whole spectrum of the host country's education and training system).

In many cases programme objectives were not achieved, particularly those relating to 'familiarisation with education and vocational training systems and occupational practice in other countries in the context of economic and political conditions'. This was because the overwhelming majority of participants used the German dual vocational training system as their reference. The evaluation suggests this was mainly due to participants not being able to put things into context indirectly demonstrating the urgent need

to support the personal development of target groups via a tailored programme such as IFKA.

Other important aspects of evaluation were the transmission of results to the appropriate specialists (in this case through publication) and the use of the experience gained in preparing exchange participants in subsequent years.

1984-93 (selected documentation of results)

From 1984 to 1993 evaluation of certain parts of IFKA concentrated on vocational training collaboration with countries regarded as significant German partners, e.g. China, Turkey, and the role model, Japan. The evaluation reports resulting from a quantitative hermeneutic text analysis related to the three selected countries referred to the previously mentioned 'traditional' programme objectives of participants' personal development. The findings also fed into country studies of China, Japan and Turkey prepared at the same time by researchers.

By 1992, more than 1 500 German IFKA participants had submitted reports on their experiences between 1983 and 1991 and the study was eventually extended to include projects up to 1994. These reports were evaluated, complemented by CDG/InWent material on the IFKA programme, CDG/InWent accountability reports to BMBW and internal CDG/InWent programme assessments. The then BMBW, CDG/InWent and DIPF cooperated on developing a design for computer-aided hermeneutic evaluation. The evaluation focused on:

- (a) professional and personal added value (including intercultural understanding) for participants;
- (b) benefit for the companies and other organisa-

Table 7: **1983-94: formative evaluation and documentation of findings**

Time period/ countries	Focuses	Type of assessment	Documented in:
Since 1983: all outgoing countries only	Organisational framework; career benefit; personal benefit; professional benefit	Computerised qualitative document analysis with database of relationally linked main categories and keywords. All participants' reports were processed accordingly. Particularly the professional aspects of the foreign vocational training systems were evaluated for the IHBB	A section of Chapter 7, on 'transnational mobility and international vocational training cooperation' in <i>Internationales Handbuch der Berufsbildung (IHBB)</i> , Baden-Baden 1995-2002

- tions employing the IFKA participants and for international vocational training cooperation;
- (c) programme organisation and implementation and programme improvement;
 - (d) transfer of results, with the following components:
 - (i) professional preparation of participants;
 - (ii) availability of results to the relevant interested specialist audience (enterprises, educational authorities, etc.), e.g. multipliers, networking, greater flexibility in the German vocational training system;
 - (iii) discussion of practical and theoretical results in the scientific community.

The evaluators concluded that the programme objectives had been achieved but that the programme could be greatly improved if changes relating to programme planning, implementation and assessment, reporting structure, training of support staff, collaboration with partner organisations abroad and dissemination of findings were introduced.

The *Internationales Handbuch der Berufsbildung* (International Vocational Training Handbook, IHBB) was and is an important medium for the circulation of the results of participants' experiences of their host countries' education and training systems. It can also be used to prepare participants in outgoing projects for their stay abroad. Today the IHBB is a loose-leaf collection

of 40 country studies and a comparative introduction totalling around 4 300 pages. Its function within IFKA has led to fierce debate in the scientific community.

The evaluation findings also provided input for their own operationalisation in programme planning, implementation and assessment. This took the form of meetings and workshops for employees of CDG/InWent and BMBF, programme participants and DIPF evaluators. The outcomes of this complex and costly formative evaluation (approximately EUR 160 000 over 18 months) were not only those visible externally, such as new reporting structures for participants and the IHBB. They also included many changes to the programme itself and to personnel training with people on the ground implementing the majority of the recommendations from the bottom up.

In 1997/98 a survey of IFKA participants from 1983 to 1996 was conducted to supplement the qualitative formative evaluation primarily based on participants' reports. In order to permit an overall assessment of IFKA, the main question of the research study, which focused essentially on people, was worded as follows: 'What impact did IFKA have on programme participants (a cause-and-effect model must be reconstructed) and on other people or institutions in some kind of contact with or in some way affected by IFKA?'

Table 8: **1983-96: summative evaluation, added value of IFKA, formative proposals for programme redesign**

Time period/countries	Focuses	Type of assessment	Documented in:
Survey of all participants from 1983 to 1996 via a questionnaire (47 % response rate)		Questionnaire data on participants from 1983 to 1996	
1995 and 1996: all outgoing countries, detailed assessments of comparative studies for Finland, Japan and the US	professional/career-oriented continuing training; international skills; international contacts; networking within the visit group.	The participants' reports stored in the database under main categories and keywords undergo second-level correlation and selection using structural content analysis (WinMax program) based on categories developed from the survey assessment.	Findings are recorded in <i>Evaluierungsstudie zum IFKA</i> .

A complex model of the conditions for a 'successful' exchange was developed on the basis of empirical analysis of available data. This created a theoretical framework for the findings. The research adopted a multimethod approach because there were several ways of accessing the matter to be evaluated and a research project goal was to deal with the question from different angles and at different levels. Various methods were therefore used to explore the facts. Since different sampling tools can produce the same or different results, the processing procedure tested the similarities and differences between the various data-gathering devices. The convergent and divergent findings of the research tools provided information on the validity of certain judgements and the possibility of performing an internal validation of the assessment procedure.

A multimethod approach was selected to examine the research questions and to compare and correlate the various types of information.

(a) Participant questionnaire data from 1983 to 1996

Former IFKA participants from 1983 to 1996 completed questionnaires on their attitudes, opinions and judgements of IFKA, using prescribed categories of answer. This made it possible to perform a correlate analysis incorporating some personal data.

(b) Participants' reports from 1995 to 1996

Since participants committed themselves to writing a report before travelling to their host country, valuable information about the visits was available immediately a visit was over. CDG/InWent provided a precise report framework to which participants were obliged to adhere. Their reports were independent, individual accounts. Utilising both data sources in conjunction allowed valid judgements to be made on the quality of implementation and on the benefits of IFKA.

In contrast to the annual evaluation which was expected to provide detailed information on both the annual programme and the individual projects and their added value for people and systems (Section 5.3.1), this summative evaluation placed the central focus on the effect taking part in IFKA had on the individual participants. The individual target countries were of interest because conditions there (organisation and programme content)

were correlated with the findings on personal benefit. This allowed conclusions to go beyond personal added value and to make recommendations for the 'ideal' IFKA project.

The questionnaire data were structured using path analysis (See Figure 4) to obtain evidence of cause and effect. Cause-and-effect analyses require longitudinal investigations to ascertain which projects have what impact (systematic variation of projects). However these seemed unlikely to occur in this case because of for example, the client's future prospects, cost, the tendering system of public bodies and the design of the study. The plethora of potential relationships was therefore reduced and the model was tested to see whether it was able to represent the available data adequately.

In order to demonstrate the potential of this evaluation procedure for key (retrospective) results and future (prospective) design changes, a number of results were quoted, with speculative recommendations.

(a) Participants' expectations from individual projects

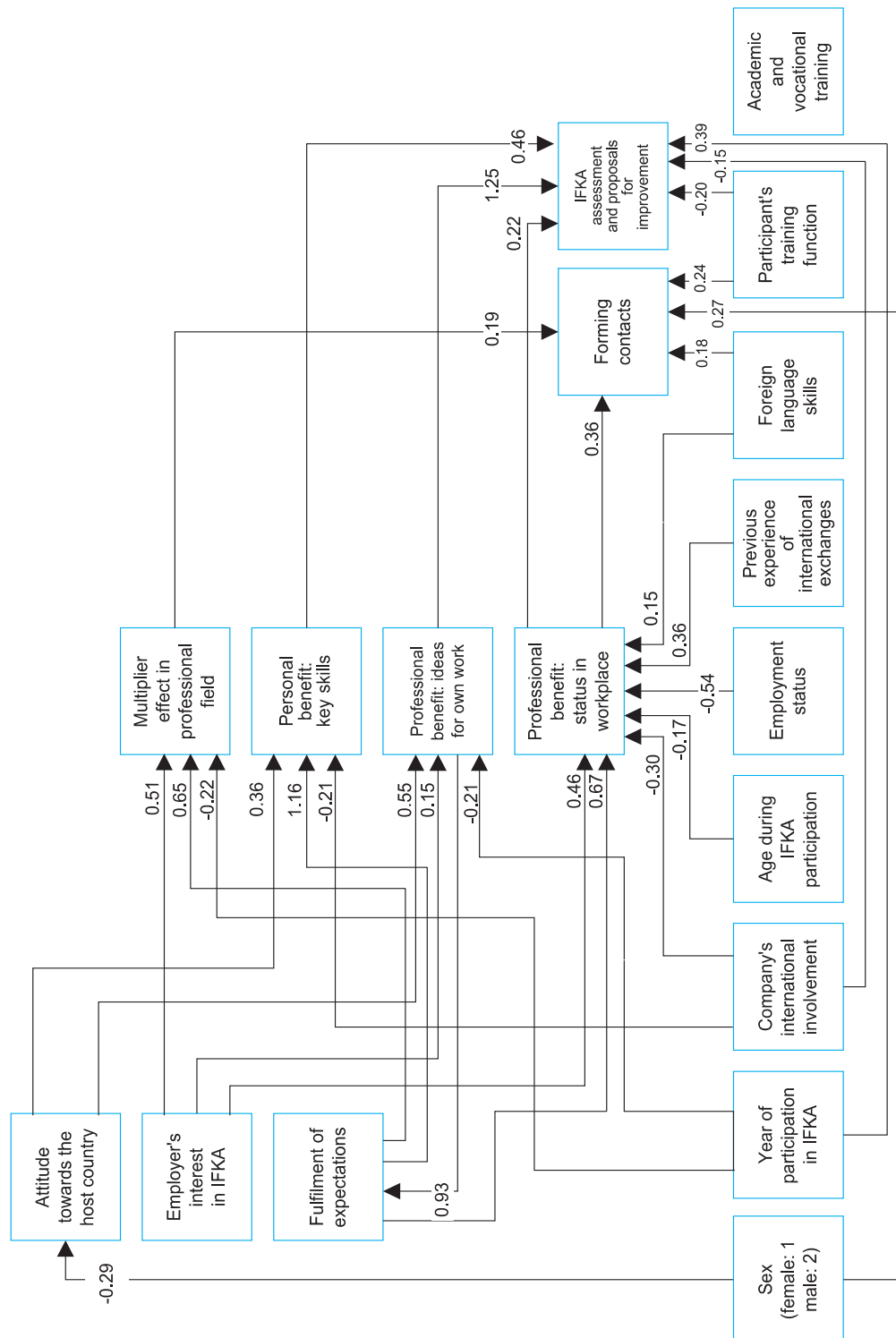
In general, participants had high expectations of their stays abroad. They mainly hoped that they would become familiar with other vocational training systems, broaden their horizons, meet people from other cultures and learn to reflect critically on their own work. The pedagogical aspect was also important to them. Participants did not expect to be able to forge links with enterprises, learn new approaches to interpersonal relationships with colleagues, improve language skills or develop friendships in such a brief period of time.

Fulfilling expectations increases the chances of participants being multipliers and obtaining key skills, thereby profiting personally. These outcomes closely related to the professional benefits of gaining ideas for participants' own work. If this aim was realised, participants were likely to assess the exchange very favourably.

(b) Personal benefit: gaining key skills

Answers to the questions demonstrate clearly that IFKA strongly inspired participants to learn about other cultures. They stated that their appreciation and understanding of the host country's culture and of aspects of daily life

Figure 4: Structural analysis of evaluation (survey of participants) 1984-96



had increased greatly as a direct result of the exchange. They also felt they were now better equipped to deal with people from other

cultures. It appears that IFKA encourages its participants to reflect more on their own culture and to be more open to foreign ideas.

- (c) Professional benefit: gaining ideas for one's own work

Taking part in exchanges gave participants new ideas for didactic, pedagogical, social and cultural aspects of their work. Visits were not as productive in exchanging and implementing specific plans or teaching and learning material. The majority of participants felt that the primary benefit of the visit was personal enrichment. The stay abroad rarely had a direct impact on participants' careers. Some reported increased status among colleagues and superiors and sometimes envy.

- (d) Contacts/networking

Information and training trips enabled many participants to make various contacts. 54 % established private contacts with people in the host country with 33 % maintaining these contacts to the present day. 26 % of participants had even taken private trips to the host country after the visit. It is therefore clear that these contacts were quite close and quite remarkable since IFKA projects are short-term and the intense schedule leaves little time for private interaction.

- (e) Employers and superiors

Employers and superiors must be actively interested in IFKA if the multiplier effect is to succeed and if participants are to gain ideas for their own work and to implement these at their place of work, fostering the development of international contacts.

- (f) Attitudes toward the host country

Positive attitudes toward the host country related to personal benefit in gaining key skills and new ideas for participants' own work.

- (g) IFKA projects affect men and women differently

The average age of women taking part in exchanges was 40, five years younger than the average male participant's age. They had better initial training qualifications than men, had superior language skills and tended to have a more positive attitude towards the host country. However, they had less experience of international exchange and usually

worked in companies with fewer foreign connections. Consequently they were far less involved in developing international contacts than men.

- (h) Opinions of IFKA improved over the years

Participants' opinions of IFKA improved over the years. However, there was a decline in personal benefit in ideas for participants' own work through exchange of training material and teaching and learning material and in the multiplier effect in motivating colleagues and trainees.

- (i) Previous experience of an international exchange improved status within the company

Previous experience of an international exchange had a positive influence on the status of IFKA participants within the company through increased competence and respect as a result of taking part in IFKA.

- (j) Academic and vocational training was not a significant determinant of success in IFKA

Analysis of the questionnaire data analysis showed that academic and vocational training had no influence on the success or failure of participation in IFKA.

- (k) Essential added value

Studies into the essential added value of IFKA revealed that its benefits were not simply derived from the directly programme-related opportunities given to participants to familiarise themselves with other vocational training systems and to gain key skills. IFKA also allows German vocational training specialists to establish contacts with enterprises and institutions in the host country and to become acquainted with them in a way that would not have been possible through mere tourist activities. Participants believed that bringing together people with similar occupational focuses created the right conditions for future commercial and project cooperation between IFKA participants and institutions and companies in the partner country. The results of the qualitative and quantitative text analysis of participants' reports from 1995 and 1996 corroborated, illustrated, refined and complemented the findings from

the questionnaire data (Table 8). The fact that the reports allowed participants to express themselves freely supports the plausibility of the statements made in the survey and the proposals for programme remodelling.

The findings of evaluations from 1983 to 1996 were published in a monograph which also contains more general chapters on exchange

research and comparative vocational training research. Supplementing this more research-oriented study, the short book 20 Jahre IFKA (Göbel et al., 1998c) targeted a wider audience and served a marketing purpose through a public relations event that supported IFKA public relations.

Figure 5: **Schematic model of 'programme success' factors, taken from surveys of participants and assessment of participants' reports**

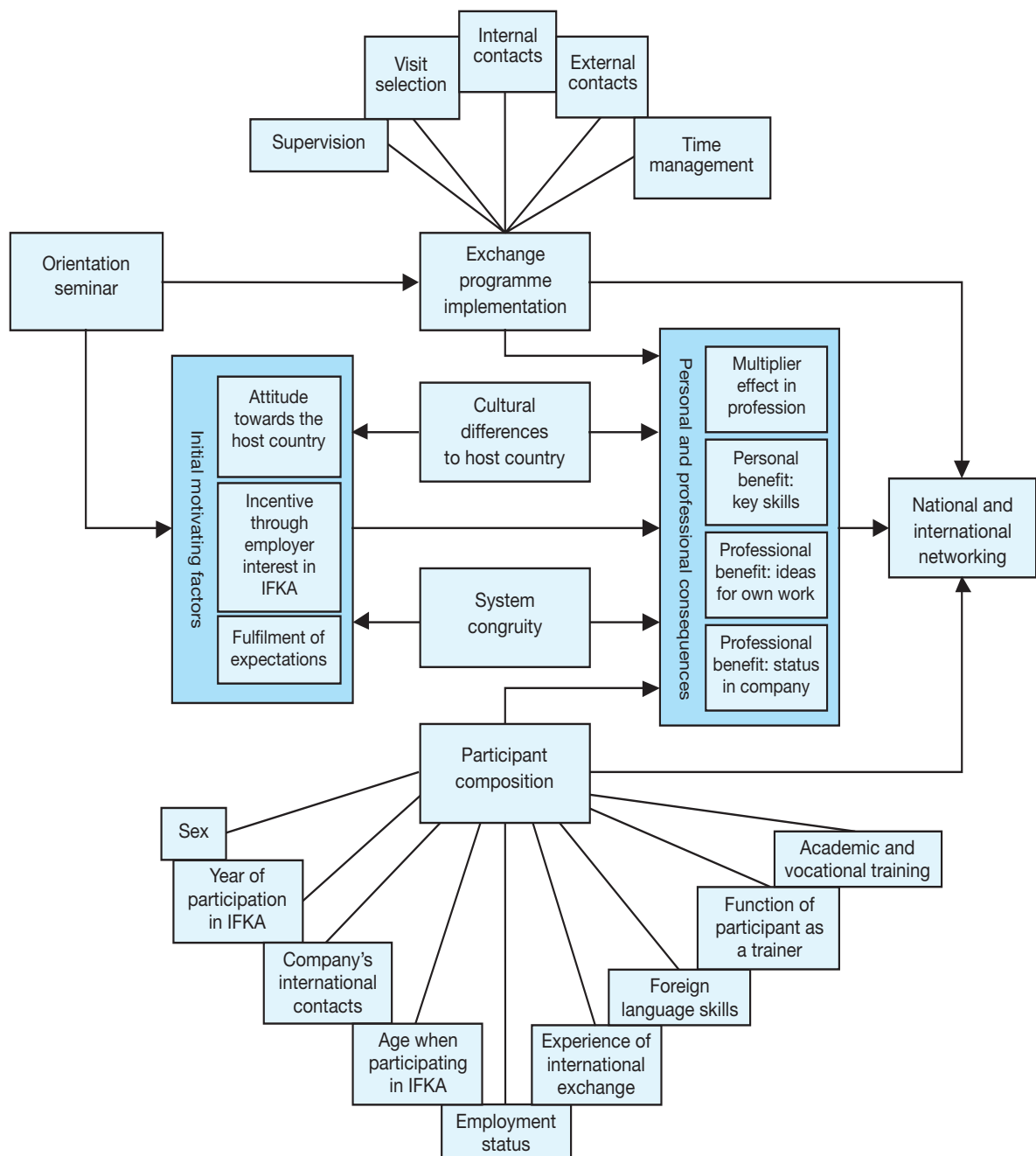


Table 9: 1995-2001: annual formative evaluation

Time period/ countries	Focuses	Type of assessment	Documented in:
1994-98: all outgoing countries only	organisational framework; career benefit; personal benefit; professional benefit multiplier function	Computerised qualitative document analysis with database. All participant reports were processed accordingly with relationally correlated main categories and keywords. Participants' reports evaluated according to all major IFKA focuses	DIPF annual report on programme evaluation for CDG/InWent and BMBF
1999-2001	organisational framework; career benefit; personal benefit; professional benefit; multiplier function; long-term effect of measures in vocational training policy (dissemination and transfer of findings) and networking by participants	Evaluation forms with closed and open answers from every participant (SPSS quantitative and qualitative assessment) Computerised qualitative document analysis with database of professional and transfer reports using the WinMaxprogram Interview with coordinators Project-specific assessments also possible because of the 78 % response rate	DIPF annual report on programme evaluation for CDG/InWent and BMBF Education policy and system dimensions of IFKA discussed in <i>Innovationen nationaler Berufsbildungssysteme von Argentinien bis Zypern</i> (Hellwig et al., 2001)

Annual evaluations have been carried out for each year from 1995 to 2001, in the year following, briefly overlapping with the 'summative evaluation 1983-96'. The last was conducted between February and August 2002.

Consequently, evaluation practice has been constantly developed and refined in the manner of formative evaluation. Advances have not only been the result of evaluators' recommendations. They have also originated in the detailed structure of the more qualitative, hermeneutic evaluations up until 1998, which were based on participants' reports on their experiences written according to the reporting framework devised in 1994. The programme organisers (CDG/InWent) and the evaluators (DIPF) maintained close communication throughout.

When vocational training policy shifted in 1999 and changes were made in IFKA's programme

objectives, a restructuring of IFKA took place, together with a number of projects, under the slogan 'IFKA 2000'. Evaluation design was an important part of the transformation. Evaluation tools such as opinion polls of participants on the organisation, or their statements on the professional added value for the German situation were jointly developed alongside the new programme structures, before programmes were implemented.

Integration of evaluation into the programme design and programme implementation allows not only the evaluation of an entire year's projects (which can be as many as 20 individual measures) but also the collection of statements on a specific outgoing project. Reporting by participants as the basis for evaluation was completely revamped. When IFKA began, the basis for evaluation was

participants' freely formulated reports on prescribed themes, the content and structure of which were often adapted to changed circumstances (see above). Later, the basis for evaluation was on professional reports concentrating on the 'added value' of adapting the German vocational training system to meet international challenges and a questionnaire with open and closed questions which participants had while projects were running.

In 1999 and 2000 these direct evaluation data were complemented by questionnaires sent to every participant a year after they had finished (e.g. evaluation of the 2000 programme included distributing questionnaires to 1999 participants) to ascertain the lasting impact of measures, particularly with regard to networking, relevance to the vocational training policy of participants' companies, and innovation transfer in the vocational training system. The programme organiser also set up an Internet platform (CDG/InWent website) to document participants' professional reports in order to support networking by participants and record innovative projects relevant to vocational training policy resulting from individual participants' encounters with a foreign vocational training system.

5.3.3. Interdisciplinary team of evaluators

Hybrid evaluation is a useful addition to qualitatively hermeneutic and quantitatively empirical evaluation methods. It established itself as the norm between 1983 and 1996, synthesising the combined efforts of research teams consisting of vocational educationists, educational psychologists and comparative VET researchers. This combination brings together academic VET skills in the country of origin and the host country with proficiencies in evaluation research and international comparative research.

Involving different scientific communities enables us to assume that evaluation research designed in interdisciplinary discussions is always current. It is, therefore, important that evaluators continue to perform active research in their own fields and do not overspecialise in the evaluation of international vocational training programmes.

5.3.4. Dissemination of findings

As discussed above, since the programme began one of its main features has been the dissemina-

tion of programme findings. In the past, participant reports were a particularly valuable source. They were the subject of several publications (Lauterbach, 1984, 1986, 1995; Henze, 1991; Georg, 1993). These standard works also served as introductory material for future participants. One major work is *Innovationen nationaler Berufsbildungssysteme von Argentinien bis Zypern* (Hellwig et al., 2001). But participants and publications are not the only disseminators of programme findings. Special events also play a vital role. They have included IFKA's 20-year jubilee in 1999 and the *International transfer and vocational qualification* conference in 2001, which involved IFKA participants, companies, national and international educational authorities, partner organisations abroad and the team of evaluators.

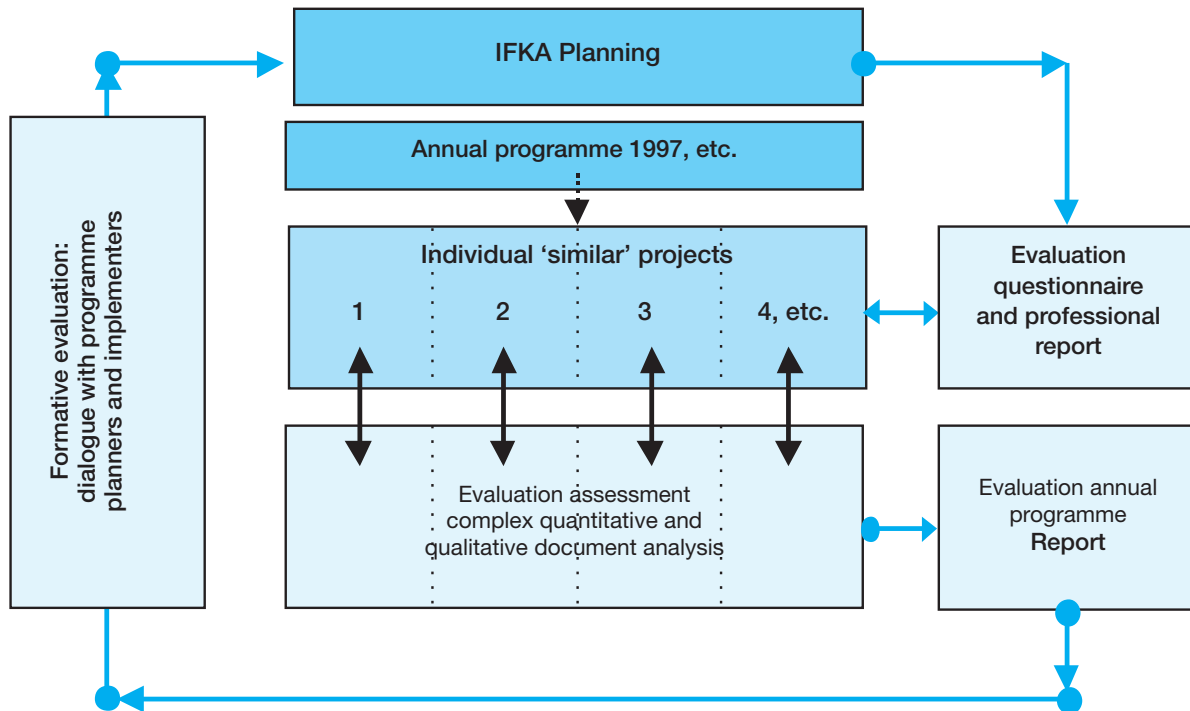
Annual evaluations have so far not considered these events or the published reports on evaluation and transfer. However, their role in making the programme better known, gaining new business partners and inspiring numerous applications for the projects should not be underestimated.

5.4. Summary

IFKA evaluations have taken place since 1984. The programme reflects the revolution in evaluation culture and the most important principles of evaluation. Methods have been constantly refined and diversified. Hermeneutic text analysis dominated initially, only to be supplanted by computer-aided text analysis complemented by participant surveys. Qualitative and quantitative social science evaluations have become an increasingly automatic part of annual evaluations and of the one evaluation of participants over 10 years carried out to date.

There is certainly also a case for making an evaluation of IFKA evaluations, given the number so far conducted and developments in evaluation techniques. The standards of scientific associations specialising in evaluation research and programme evaluation would lend themselves to this endeavour. As mentioned above, the Joint Committee on Standards for Educational Evaluation (JCSEE) and the German Evaluation Society (DeGEval) have a series of evaluation criteria: (1) utility, (2) feasibility, (3) fairness and (4) accuracy.

Figure 6: IFKA evaluation structures



The utility of IFKA evaluations is undisputed in the context of increased integration of evaluation tools into the operation of the programme, changes in programme structures and programme outcomes such as conferences, publications and transfer of innovations to the German system. Given the use made of the evaluation tools in ascertaining 'true' project and programme outcomes, their utility in transferring innovations and further developing the programme is also a foregone conclusion. The integration of data collection into project implementation increases the 'feasibility' of evaluation

because at the end of each measure most of the anticipated material is available for assessment (except the results of surveys conducted one year later). Since feasibility is hardly possible without the stakeholders (organisers and participants), the aim is to achieve a permanent dialogue. To ensure 'fairness', all stakeholders should have access wherever possible to evaluation results. Some shortcomings in this area have yet to be remedied. The programme's reliance on scientific methods is a step in the direction of 'accuracy'. Meta-evaluation would be of advantage.

6. Summary and recommendations for future practice

6.1. Four practical models

Selection of the four practical models was not limited to evaluations of EU mobility programmes because practical evaluation began with the ‘older’ international exchange and promotion programmes of other providers, such as individual countries or foundations and more experience was available in such fields. Two of the examples considered (DFJW and IFKA) have

been part of evaluation practice for years and are a match for the DeGEval standards. Both of them broadly correspond in design to the examples of EU programmes. DFJW is to be considered comparable to LdV in its variety and complexity, whilst IFKA and CSV are similar in the clear structure that they both exhibit.

The availability of documentation on results, access to appraisals and interview opportunities were also central to the selection of examples.

Table 10: **Mobility in VET – one of many ways of moving towards the principal objective**

DFJW	LdV
The various DFJW projects are designed to further the principal priority, the promotion of international understanding. VET furnishes one means of communication, so that over the course of the years exchange activities have evolved in this field. These activities differ greatly (i.e. apprentice exchanges, work experience for young people in initial training, exchanges between vocational schools, exchanges of trainers, long-term and short-term measures).	The LdV Programme was created to promote better mutual understanding of VET in the EU countries. Mobility is one of the LdV programme areas. The various measures in this field serve the above-mentioned principal aim of LdV, but their forms may differ quite substantially (i.e. apprentice exchange, work experience programmes for young people in initial training, exchanges between vocational schools, exchanges of trainers, long-term and short-term measures, CSV).

Table 11: **Mobility in VET – two concrete measures**

IFKA	CSV
IFKA set itself the goal of promoting VET in Germany by means of 14-day in-service training visits abroad that focus on particular themes to give specific target groups the opportunity of individual continuing training and to encourage innovation in VET in Germany. New national and international networks are created through the process of group dynamics.	CSV assists vocational training decision-makers from Member States by organising group visits to EU countries lasting three to five days and focusing on particular themes. The intention is to create a better understanding of other VET systems (or aspects of systems and political decisions) so that what is learnt can be used by participants in their own working environment in the European context. CSV supports construction of international networks.

6.2. Implementation of evaluation and practical examples

Programme designers and organisers and funders of programmes, particularly those oper-

ating at a political level, want programme evaluation to provide feedback and explain and justify expenditure of public funds. They also hope for perpetual improvement of actual programmes by means of more precisely defined goals. Are these

expectations fulfilled by the four cases mentioned above? The following results provide an interim summary and combine theoretical and practical aspects of the cases. The order in which points (1) to (15) are placed refers to the order of an evaluation based on the evaluation standards discussed in Section 1.3.

The evaluation team began with:

- (a) programme organisation, in order to examine the focus, structure and organisation of the programme;
- (b) programme goals were the next stage, concentrating on evaluable operationalisation; subsequent comments referred to:
- (c) quality of evaluation. Success is not solely based on the methodological stringency of an evaluation design in a narrow sense, but relates to:
- (d) transformation of evaluation results into the practice of programme implementation, an essential criterion for 'successful', 'exemplary' evaluation.

6.2.1. Programme organisation, evaluation planning and preparation

1. Programme content and structure

A meaningful and practice-oriented programme evaluation that draws on the most recent scientific standards requires a thorough knowledge of the programme content and an appropriate analysis of programme structure. The structure of the individual elements of the programme should be taken into account. Determination of the suitable method for the evaluation should be based on programme structure and the programme budget and may depend on the budget for activities. Consideration of suitable methods of evaluation during the design phase of a programme significantly heightens the effectiveness of formative and summative evaluation. For this reason evaluation specialists should be involved in the design phase of VET mobility programmes.

2. Highly skilled interdisciplinary research teams

The strict demands of evaluation require highly skilled personnel. A combination of such skills in one individual is extremely rare. In the programmes included here, a realistic approach would therefore be cooperation

between professional researchers, educational psychologists and comparative VET researchers within interdisciplinary research teams. This combination unites academic qualifications related to VET in the home and host country or countries with skills in evaluation research and international comparative research. This involvement in a discipline enables researchers to keep pace with their 'own' discipline.

Specialisation in evaluating international VET programmes without this firm basis will not meet the high expectations made of evaluations. Qualitative hermeneutic and quantitative empirical methods can be complemented by these interdisciplinary competences. In the IFKA case study this combination of skills was applied to the evaluation.

3. Harmonisation of evaluation goals, content and tasks between players

Prior to an evaluation, the basic goals of quality assurance measures should be agreed between those involved. Frequently there is substantial disagreement on the application of evaluation findings. Programme designers chiefly expect proposals for improvements in implementation practice, whilst evaluators (and frequently funding bodies) are more concerned with the individual and social impact of programmes.

A common understanding of evaluation goals, contents and tasks is an essential precondition if those involved wish to work together to improve quality.

4. International evaluation projects and potential misunderstandings

The analysis of DFJW and LdV shows that varied goal interpretation and evaluation paradigms are to be expected in the case of differentiated international evaluation projects. The risk of potential misunderstandings increases with growing interdisciplinary and international involvement. Precise definition of central concepts, procedures and analytic methods is necessary for this reason. Operationalisation presupposes definition and limitation of concepts. In German VET alone, the literature offers a total of 654 variant definitions of 'key skills' during a 10-year span

(Weinert, 2000). Evaluation can, therefore, help to clarify research concepts.

5. Linking methods and concepts of evaluation

In the context of 'skills for Europe' (DFJW), Demorgon et al. draw attention to the close meshing of tools and concepts in evaluation. They even go so far as to say that it is improper to transfer evaluation strategies developed in other fields and that it is better practice to develop separate procedures for each individual topic. This amounts to a thoroughgoing constructivist approach. Irrespective of whether one shares this opinion, it does demonstrate the need for tightly defined concepts and evaluation methods. To date, this approach has seldom been applied in evaluation practice. It is cumbersome and, although it promises excellent results, it is certainly more 'expensive' than a 'standard evaluation' applied to all manner of cases, which can be restricted to counting survey figures.

6.2.2. Programme goals, development of suitable methods and tools for evaluation

6. Operationalisation of programme goals

Programme goals are usually expressed in very general, abstract and broad terms. It is essential that they be operationalised, i.e. reformulated in a less abstract manner so that their evaluatory basis consists of observable and, where possible, measurable phenomena. This cannot be achieved without selecting and prioritising variables. To eliminate arbitrariness in the reduction and operationalisation process, it is necessary to define the precise programme goals in consultation with the client during the design phase, possibly in a multistage procedure which includes provision for subsequent correction and refinement. The role of the teams of evaluators consists of advising programme administrators, in the light of thorough methodological and theory-based expertise, particularly on the effects of decisions made during the design phase on the effectiveness of evaluation projects. Tools that are to be used or developed for a given evaluation must be sensitive enough to register and reflect

desired changes. The role of the client is to monitor the feasibility and viability of the evaluation.

7. Evaluation of the long-term impact of programme goals

Programme administrators' stated goals go beyond mere concepts and include very broad contextual and causal relations affecting the effectiveness of measures. However, it is seldom possible to include complex contexts and long-term processes in evaluation. Consequently, individual evaluation projects need to refer to available findings from theoretical and empirical research to fill, by interpolation, gaps in explanations or forecasts not encompassed in a given time frame or context and thereby to offer at least some hypothetical conclusions. Another option is to connect series of evaluations in a programme. This requires not only access to previous evaluations and their data, but also presupposes their compatibility. It is then possible to verify the stability of old findings and to show medium and long-term causal chains. It would be better in many respects (cost, results, benefit for the programme) if long-term evaluation were incorporated in programme planning and annual operationalisation. Moreover, assessment can be planned in longer phases, e.g. every three years. The duration of phases should be governed by the parameters of the programme (e.g. duration and sustainability of programme goals).

6.2.3. Quality of evaluation, optimisation potential

8. Relationship between evaluation design and evaluation results

Because of the extreme complexity and general lack of feasibility of control group design, there are always several alternative explanations and prognoses which should not be accepted solely on the strength of one evaluation. The implications are therefore always limited and related to the scope of an evaluation design. This is especially true of evaluations whose results are expressed vaguely. It is therefore wise to regard the results of an evaluation as hypotheses which should feed into further evaluation.

9. Match between an overall programme and individual programmes and measures

The study of the series of mobility measures in the LdV programme has revealed many ways in which evaluation may be optimised. Although most of the elements of programmes and measures are covered by evaluations conducted by the three implementation centres, InWent, ZAV and DAAD, there is a need for coordination between individual steps and evaluation of the entire programme in order to avoid repetition and to improve the validity of small steps that build one on the other.

10. Long-term impact of programme goals

The inability of the majority of evaluations to create a control group is a particular problem. Three possibilities may be imported from scientific research methodology:

- statistical control;
- complex models;
- internal comparison applying modern qualitative evaluation strategies.

These possibilities all have the same essential purpose: the essence is to analyse the fluctuation of key variables to monitor differences in effect. Statistical control aims to concentrate on pertinent sets of variables by monitoring and systematically segregating 'disruptive' factors. Complex models such as path analysis (see IFKA) include all variables which may be regarded as disruptors or moderators of effects and make them calculable. The first two possibilities mentioned above assume a quantitative method. The same purpose may be served by a qualitative approach. Factors potentially responsible for the success or failure of programmes can be identified by analytic separation of successful and less successful programme components or projects.

11. Meaningful evaluation results from several aggregate levels and continuity of evaluation

Programmes are evaluated at many aggregate levels. The success or failure of a given project as part of a programme contributes to the success or failure of the programme at a national or international level. Empirically

based statements which include data from the level of the individual participant up to that of the system require highly complex evaluation designs and analysis procedures which surpass usual EU-level VET evaluation projects in their scope and resource requirements. Such innovative evaluations by TIMSS, PISA and the forthcoming DESI project are based on expensive designs incorporating several levels in the evaluation⁽²⁶⁾. They aim to make statements on how effects at individual (student) level relate to effects at system (school) level. Quantitative multilevel analysis has shown that there can be surprises: causes of positive effects at an individual level will not necessarily have such effects on a system level and vice versa. More such level-specific effects are to be expected when investigation involves large numbers of levels as are found in surveys at a supranational level. It is therefore recommended that representative large-scale assessments which incorporate several aggregate levels should at least occasionally be conducted within the European framework too. All individual evaluation projects should be designed to assure data compatibility. It seems reasonable also to have regard to continuity in evaluation in order to guarantee progress in evaluation practice and programme optimisation.

12. Design and methods of evaluation projects in VET in the light of the OECD school performance study

The advanced level of design and methodology in the OECD school performance studies has set standards for further evaluations in education. For the first time the representativeness of participants was a priority and probability models which use state of the art survey methods were applied. The fruits of this model survey are to be seen in the most recent evaluation application (DESI), which features new developments such as the inclusion of large-scale assessment of intercultural competence⁽²⁷⁾. There is no reason why such evaluation projects should not also be conducted at a high methodological level in VET.

⁽²⁶⁾ In the case of DESI, 11 000 pupils and 200 schools are representative of all *Länder* in the German Federal Republic.

⁽²⁷⁾ DIPF acts as the secretariat for the Deutsch-Englisch-Studie-International project (DESI). The project module to develop tools for intercultural competence is headed by Kerstin Göbel and Hermann-Günter Hesse.

6.2.4. Transformation of evaluation results and financial framework

13. Top-down or bottom-up evaluation?

It is possible to differentiate between two aspects: the treatment of multiple programmes such as LdV and DFJW, which are split into a series of individual programmes, and hermetic individual programmes. Individual programmes (e.g. CSV as a component of LdV) may, however, conceal many individual measures. LdV and DFJW evaluate individual programmes all of which together constitute the overall programme. Evaluation should be done at the individual programme level (bottom up) because of the variety of programmes alone. As the example of LdV implementation in Germany shows, there is much worthwhile evaluation, but no agreed design for formative evaluation has yet been developed and agreed by the implementing agencies. Similarly, top-down evaluation still dominates DFJW, where this process is still apparently in favour because of the established evaluation culture.

Transparent programmes such as CSV and IFKA are also characterised by bottom-up and top-down effects. When external and political arguments are to the fore, as in the case of the CSV evaluation 2000/2001, there is less emphasis on programme organisers (Cedefop, NLOs) and participants. The abstract results of these top-down evaluations make them more akin to consultation and political justification. The bottom-up effect, advantageous to the long-term development of programmes, is replaced by a top-down effect. Such summative evaluation addresses neither participants nor programme organisers in the realisation of programme goals via individual measures. Since evaluation results provide the input for operationalised programme planning, implementation and appraisal, which can only be successfully achieved with the cooperation of organisers and programme participants, the integration of these stakeholders in design and operation is a significant bottom-up factor in the effectiveness (intervention potential) of evaluations. Such a bottom-up procedure has been adopted in CSV for participants' reports.

14. Dissemination of results

The dissemination of results is of great importance for the success of an evaluation. Evaluators of VET international exchange programmes often consider their duties discharged once data has been collated. The systematic dissemination and discussion of results do not always receive due consideration. Equally, public debate ensuing from a broad dissemination of the stages and results of evaluation is rare. Transparency and time for discussion of the evaluation are, however, much more important for those directly affected and, in particular, implementation centres. This is true of both formative and summative evaluation. Electronic media are particularly useful in such cases. In order to achieve the intended broad effect, target groups should also be included in dissemination.

Apart from these internal optimisation considerations, the dissemination of scientific findings and experiences among a broad public provides a platform that may benefit future evaluations.

The funding arrangements for an evaluation project should allocate sufficient resources for the transfer of results. The production of evaluation studies simply destined to gather dust at the back of a shelf is not desirable and the fact that a good plan for the dissemination of results (in printed, electronic and/or conference form) will incur costs is often not considered.

15. Financial framework, evaluation concept and political mediation

Programme planners and funders, particularly those operating at a political level, want programme evaluation (a) to justify the expenditure of public funds and possibly (b) to optimise programmes by means of precisely defined goals. The availability of funding has an indirect effect on the scale of an evaluation and thus on evaluation procedure. To date, (a) unequivocally dominates as it is easier to justify politically to the public. A complex 'intervention' to ensure programme optimisation, based on evaluation results (b) is less convincing within such a context and also more costly. Equally, funding for evaluation designs which must clearly be regarded as falling under (b)

must be secured by programme organisers through persistent political pressure and by appropriating arguments largely from (a) (see IFKA). The consequences of 'optimising' a programme, such as lower organisational costs, satisfaction of participants, easier marketing or durability of programme goals achieved, seem to be of little significance in the allocation of resources. If, however, the public feels politically 'alarmed,' as is the case in large-scale assessments of school performance such as PISA, then financial support for evaluation is more readily forthcoming. These results lead in turn to a direct demand for greater dissemination of programme results to the politically aware public.

6.3. Prospects for a new evaluation culture and 12 proposals for the implementation of evaluation

The four case studies presented here demonstrate very different approaches to the evaluation of their programmes. The continual development of evaluation methods in EU mobility programmes stands out. Evaluation has obviously received more attention and thus greater financial support so that, given the continuation of this trend, we can expect a continual rise in the quality of evaluation of VET exchanges within the EU.

The other two programmes presented here (IFKA and DFJW) were pioneers in the design and content of evaluation. Today they are at risk of losing substantial elements due to budget cuts. IFKA will only be evaluated every three years from 2003, which will prevent data generation immediately during or after an exchange (by InWEnt GmbH), hinder the continuous advising of programme organisers by evaluators and call into question the inherent validity of formative evaluation. There will, therefore, no longer be any basis for re-interpretation through evaluation. The evaluation practice of DFJW is equally subject to the funding provided by sponsors. Vacillation in this respect precludes longitudinal investigation.

Today, evaluations are complex processes less dependent on political justification than on the dialogue between stakeholders such as funders, implementation centres, providers and partici-

pants. Measuring programme output and its documentation is a contribution to programme design which has internal consequences for programme organisation and its tools and external consequences for political practice.

Evaluation is therefore a public affair rather than an activity of government monitoring agencies such as an auditing commission. The dissemination of results to the (expert) public is an important instrument in creating transparency and in a public culture of evaluation. These results could also lead to evaluations no longer being motivated by political justification but to greater emphasis on the linking of evaluation and intervention in programme optimisation. This new situation would also mean that the costs of evaluation were regarded as an integral part of programme costs.

Developing such an evaluation culture presupposes a minimum of mutual confidence between stakeholders. 'Fear' of a bad report and an apologist reaction to critical evaluation findings are counterproductive to developing this new, inclusive evaluation culture. Everyone can pursue the common goal of programme implementation and optimisation. Furthermore, those involved expect this from evaluation. Only where a stakeholder or a group demonstrably reacts negatively should the question of their replacement be entertained.

Complementary to these basic remarks on evaluation culture, this report concludes with twelve brief general proposals containing essential rules for the conduct of evaluation in international exchange programmes and how evaluation can support the successful implementation of a programme.

Proposal 1

The measures in a programme should be organised to enable formative evaluation making evaluation an integral part of programme implementation. 'Formative' means acquiring knowledge on how to optimise programmes, i.e., on how to plan and implement interventions and disseminate knowledge and information to support the enhancement of programme outcomes in line with goals.

Proposal 2

Evaluation and scientific research should be interrelated. Scientific work means hypothesis-based and quasi-experimental work. It supports the acquisition of knowledge on the relativity of impact factors.

Proposal 3

No evaluation project can stand alone. Each evaluation project must refer to knowledge outside the project, that is, to theory. The programmes to be evaluated are highly complex. They have multiple goals, not all of which harmonise. They refer to different levels, individual, institutional, systemic and cultural, and they offer different perspectives from different groups of stakeholders. The consequences are: the same goal can be accessed in different ways whilst the same method may lead to different outcomes with distinct effects and thus to possible conflict situations, with one goal achieved at the cost of another. Fundamental disagreements over the possible scale of the impact due to different theoretical foundations are a good reason for clarification of programme strategies. The theory-driven approach is motivated by the need to bridge the gap between description and the need for elucidation in order to optimise the programme.

Proposal 4

Evaluation should be useful for stakeholders. Evaluation must be designed so as to respond to stakeholders' questions about the matter to be evaluated. Because of the normative character of evaluation, clarification, overt explication and refinement (operationalisation) are a necessary first step in the implementation of a programme. This aids precise definition of the goals, measures and expectations of a programme and, ultimately, its practicability.

Proposal 5

The theory-driven approach is the basis of generalisation which draws on the results of evaluations of different programmes. Evaluation, if theoretically based, thus supports the accumulation of programme implementation knowledge. How can evaluation be used before, during and after a programme or revised implementation to feed back into the reform process or to alter the characteristics of a programme or a reform which do not appear (any longer) appropriate?

Proposal 6

The shift in priority from 'objective' criteria (reliability, validity, objectivity) to 'constructive/constructivist' criteria (communication, intervention, transparency, relevance, developing alternatives) is concomitant with an increase in the applicability of evaluation. (There need not be a contradiction between objec-

tive and constructive criteria.) All 'new' criteria involve the interaction of all participants/stakeholders. Evaluation is no longer seen as an external approach to the programme being evaluated. Evaluators are instead experts instrumental to the programme. Evaluation is thus a continuous interactive process which can be adapted during all phases of programme implementation.

Proposal 7

The paths which lead to a product's evaluation can be reconstructed by way of an inductive strategy that should be grounded on theoretical foundations. This strategy includes the clarification of concepts and methods. Evaluations planned in this fashion can contribute to the accumulation of knowledge that is useful not only for the programme in question but also for future programmes.

Proposal 8

Before the implementation of a programme, it is important to estimate the intended scale of the impact. Measures and evaluation instruments have to be selected and developed with this in mind. If they are not sensitive enough to record outcomes, they will furnish no information on impact.

Proposal 9

Before the programme is implemented, complexity, as reflected by the scale of the impact, should be taken into account, not only the intended effects but also side effects that could alter or even prevent the success of the outcome. What conditions must be fulfilled to ensure that evaluation is a learning exercise for all those involved, including policy-makers?

Proposal 10

Evaluators should be experts in their specific field, with methodological and strategic knowledge to ensure the usefulness of the evaluation process for those concerned.

Proposal 11

Evaluations should be designed so as to guarantee interconnectability with other evaluations.

Proposal 12

The practical realisation of evaluation results is to be regarded as a step in its own right. Intervention in the form of 'action research' should be programmed and evaluated in an appropriate manner.

List of abbreviations

BA	<i>Bundesanstalt für Arbeit</i> / Federal Employment service
BIBB	<i>Bundesinstitut für Berufsbildung</i> / Federal institute for VET
BMBF	<i>Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie</i> / Federal Ministry of Education, Science, Research and Technology (from 1995)
BMBF	<i>Bundesministerium für Bildung und Forschung</i> / Federal Ministry of Education and Research (from 1998)
BMBW	<i>Bundesministerium für Bildung und Wissenschaft</i> / Federal Ministry of Education and Science (to 1994)
CDG	<i>Carl Duisberg Gesellschaft</i> / Carl Duisberg Society
Cedefop	<i>Centre européen pour le développement de la formation professionnelle</i> / European Centre for the Development of Vocational Training
Comett	EU academic and business technology cooperation programme
CSV	Community Study Visits Programme for those responsible for vocational training
DAAD	<i>Deutscher Akademischer Austauschdienst</i> / German academic exchange service
DeGEval	<i>Deutsche Gesellschaft für Evaluation e.V.</i> / German Evaluation Society
DESI	<i>Deutsch-Englisch-Schülerleistungen-International</i> / German-English Pupil Performance International
DFJW	<i>Deutsch-Französisches-Jugendwerk</i> /German-French youth foundation
DIPF	<i>Deutsches Institut für Internationale Pädagogische Forschung</i> / German Institute for international educational research
DSE	<i>Deutsche Stiftung für Internationale Entwicklung</i> / German foundation for international development
EEA	European economic area
EFTA	European free trade area
ESF	European social fund
ETF	European Training Foundation
Eurotecnet	Initiative promoting vocational training innovation in view of the technological transformation in the EU.
FORCE	<i>Formation continue en Europe</i> . European Community initiative to promote vocational training among young people in Member States and to prepare them for adult and working life.
GTZ	<i>Deutsche Gesellschaft für Technische Zusammenarbeit</i> / German Society for technical cooperation
IALS	International adult literacy survey
IBE	International bureau of education
IFKA	<i>Internationaler Fachkräfteaustausch (in der Berufsbildung)</i> / International exchange of VET specialists
IHBB	<i>Internationales Handbuch der Berufsbildung</i> / International handbook of VET
ILO	International Labour Organisation
InWent	<i>Internationale Weiterbildung und Entwicklung gemeinnützige GmbH</i> / International continuing education and development non-profit Ltd.

JCSEE	Joint committee on standards for educational evaluation
LdV	Leonardo da Vinci (programme) Initiative for introducing EU vocational training policy
Lingua	European Community programme promoting language teaching and learning
NLO	National liaison officer
OECD	Organisation for Economic Cooperation and Development
PETRA	Partnership in education and training European Community initiative to promote vocational training among young people in Member States and to prepare them for adult and working life
PHARE	Poland and Hungary, aid for restructuring of the economies
PISA	Programme for international student assessment
SME	Small and medium-sized enterprise
TIMSS	Third international mathematics and science study
TVET	Technical and vocational education and training
Unesco	United Nations educational, scientific and cultural organisation
VET	Vocational education and training
VmQ	<i>Verfahren modularer Qualifizierung</i> / modular qualification system
WSF	<i>Institut Wirtschaft- und Sozialforschung</i> (Economic and social research institute)
ZAV	<i>Zentralstelle für Arbeitsvermittlung</i> (Central placement office)

References and bibliography

- BIBB. Valorisation durch Evaluation. Ergebnisse europäischer Innovationsprojekte in Leonardo da Vinci I (1995-1999). *Impuls (Leonardo da Vinci)*, 2001, Vol. 1.
- Busse, G.; Fahle, K. *Lernen durch Mobilität. Untersuchung zu längerfristigen Auslandsaufenthalten von Auszubildenden und jungen Berufstätigen im Auftrag von Cedefop*. Dortmund: CDG – Carl Duisberg Society, 1999.
- Campbell, D. T.; Fiske, D. W. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 1959, Vol. 56, p. 81-105.
- CDG – Carl Duisberg Gesellschaft. *Kompendium 2001: Leonardo da Vinci: Vermittlungs- und Austauschmassnahmen der Antragsrunde 2001 [Compendium 2001: Leonardo da Vinci: placement and exchange measures from the 2001 round of applications]* / Bundesanstalt für Arbeit – BA. Bonn: CDG, 2001.
- CDG. *Berufliche Weiterbildung im Ausland: Programme 2001*. Cologne: CDG – Carl Duisberg Society, 2001.
- CDG; Federal Ministry of Education and Research, brochures, manuscripts, etc. (not published). Bonn, Cologne: 1997 ff.
- CDG; Federal Ministry of Education and Research. *Berufsbildung im weltweiten Dialog. 20 Jahre Internationaler Fachkräfteaustausch in der Berufsbildung*. Cologne: CDG – Carl Duisberg Society, 1998.
- CDG; Federal Ministry of Education and Research. *Internationaler Innovationstransfer und berufliche Qualifizierung am Beispiel des Internationalen Fachkräfteaustausches in der Berufsbildung (IFKA)*. Berlin, 18–19 October 2001. Documentation. Cologne: CDG – Carl Duisberg Society, 2002.
- CDG; Federal Ministry of Education and Research. *Weiterbildung ohne Grenzen 2002. Angebote zur beruflichen Qualifizierung im Ausland*. Cologne: CDG – Carl Duisberg Society, 1959.
- Cedefop. *Annual report 2001*. Luxembourg: Office for Official Publications of the European Communities, 2002b.
- Cedefop. *Katalog – Catalogue 2003*. Study Visits. Thessaloniki: Cedefop, 2002a.
- Chen, H.; Rossi, P. H. Evaluating with sense: the theory-driven approach. In: *Evaluation Review*, 1983, Vol. 7, pp. 283-302.
- DeGEval – German Evaluation Society. Beywl, W. (ed.) *Standards für Evaluation*. Cologne: DeGEval, 2002. Available from Internet: <http://www.degeval.de/standards/standards.htm> [cited 11.12.2003].
- Demorgon, J.; et al. *Europakompetenz lernen*. Frankfurt: Campus, 2001.
- DFJW – Deutsch-Französisches-Jugendwerk (ed.) *Activity report 1993*. Bad Honnef: DFJW, 1994.
- DFJW – Deutsch-Französisches-Jugendwerk (ed.) *Activity report 2001*. Bad Honnef, Paris: DFJW, 2002a.
- DFJW – Deutsch-Französisches-Jugendwerk (ed.) *Richtlinien DFJW/ Directives OFAJ*. Bad Honnef: DFJW, 2002b (Version of 1.1.2002).
- EMNID. *Die 15–24 jährigen und der deutsch-französische Jugendaustausch: Repräsentativumfrage im Auftr. d. Dt.-Franz. Jugendwerks* (DFJW), October 1985; Documentation = *Les 15-24 ans et les échanges franco-allemands de jeunes*. Bielefeld, Paris: 1985.
- Europabüro für Projektbegleitung GmbH. *Gemeinschaftsinitiative Beschäftigung. efp-Leitfaden Projektevaluation. efp-news 7*. Special edition. European Community National Coordination Office. Bonn: 1999.
- European Commission, Directorate-General for Education and Culture. *External evaluation of the European Centre for the Development of Vocational Training – Cedefop. Final Report*. Brussels: European Commission, 2001.
- European Commission. *Abschlussbericht über die Durchführung der ersten Phase des Gemeinschaftlichen Aktionsprogramms Leonardo da Vinci (1995-1999)*. Luxembourg: Office for Official Publications of the European Communities, 2000 (COM(2000)863 final).
- Fahle, K. Die Bedeutung von Auslandserfahrungen. Findings of the Leonardo da Vinci annual meeting on mobility, 24-25 September 2001. In: *Impuls (Leonardo da Vinci)*, Vol. 2, 2002.

- Friedrich, W.; Schumacher, 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.
- Georg, W. *Berufliche Bildung des Auslands: Japan. Zum Zusammenhang von Qualifizierung und Beschäftigung in Japan im Vergleich zur Bundesrepublik Deutschland*. Baden-Baden: Nomos Verlag. (Carl Duisberg Society series: Internationale Weiterbildung, Austausch, Entwicklung, Vol. 8).
- Göbel, K.; Hesse, H. G.; Lauterbach, U. *20 Jahre IFKA*. Cologne: Carl Duisberg Gesellschaft, 1998c.
- Göbel, K.; Hesse, H. G.; Lauterbach, U. *Berufsbildung im europäischen Dialog. Studie über das EU-Studienprogramm: Ergebnisse und Perspektiven*. Cologne: Carl Duisberg Gesellschaft, 1999.
- Göbel, K.; Hesse, H. G.; Lauterbach, U. Die berufliche Bildung braucht den Dialog mit dem Ausland: der internationale Fachkräfteaustausch (IFKA) und seine Folgewirkungen. Research study. Baden-Baden: Nomos Verlag, 1998a (Carl Duisberg Society series: Internationale Weiterbildung, Austausch, Entwicklung, Vol. 10).
- Göbel, K.; Hesse, H. G.; Lauterbach, U. *Kurz-Evaluierung des EU-Studienbesuchsprogramms (EUS)*. Frankfurt: DIPF, 1998b.
- Hager, W. Zur Wirksamkeit von Interventionsprogrammen: Allgemeine Kriterien der Wirksamkeit von Programmen in einzelnen Untersuchungen. In: Hager, W.; Patry, J.-L.; Brezing, H. (eds) *Evaluation psychologischer Interventionsmaßnahmen. Standards und Kriterien: Ein Handbuch*. Bern: Hans Huber, 2000, pp. 153-168.
- Hager, W.; Patry, J.-L.; Brezing, H. (eds) *Evaluation psychologischer Interventionsmaßnahmen. Standards und Kriterien: Ein Handbuch*. Bern: Hans Huber, 2000.
- Hegenbarth, G.; Stamm-Riemer, I.; Umbesen, P. *Die Zukunft beginnt mit dem Lernen von heute. Ein Leitfaden zur Qualität transnationaler Austauschpraxis*. Berlin: Understandingbus, 2001.
- Hellwig, W.; Lauterbach, U.; von Kopp, B. (eds) *Innovationen nationaler Berufsbildungssysteme von Argentinien bis Zypern*. Baden-Baden: Nomos Verlag, 2001 (Carl Duisberg Society series, Vol. 11).
- Henze, J. *Berufliche Bildung des Auslands: Volkssrepublik China*. Baden-Baden: Nomos Verlag, 1993 (Carl Duisberg Society series: Internationale Weiterbildung, Austausch, Entwicklung, Vol. 7).
- ITB – Institut Technik und Bildung. Die Programmevaluation. Manual. Der Programmträger. BLK-Programm: Neue Lernkonzepte in der dualen Berufsausbildung. Bremen: ITB Institut Technik und Bildung, 2001.
- JCSEE – Joint Committee on Standards for Educational Evaluation (ed.) *Handbuch der Evaluationsstandards* (Second edition). Opladen: Leske and Budrich, 2000.
- JCSEE – Joint Committee on Standards for Educational Evaluation. *Standards for evaluations of educational programs, projects, and materials*. New York: McGraw, 1994.
- Jöreskog, K. G.; Sörbom, D. *Advances in factor analysis and structural equation models*. Cambridge, MA: Abt Books, 1979.
- Kohler, A. *Quality in mobility. Quality award for Austrian Leonardo da Vinci mobility projects*. Presentation at Agora 13, Mobility as a didactic tool, 29-30 October 2001 in Thessaloniki.
- Lauterbach, U. *Berufliche Bildung des Auslands aus der Sicht von Ausbildern*. Carl Duisberg Society (ed.). Stuttgart: Echo-Verlag, 1984.
- Lauterbach, U. *Berufliche Bildung des Auslands: Türkei. Mit einem Vergleich der Gesetzgebung zur beruflichen Bildung*. Baden-Baden: Nomos Verlag, 1993 (Carl Duisberg Society series: Internationale Weiterbildung, Austausch, Entwicklung, Vol. 3).
- Lauterbach, U. Die Carl Duisberg Gesellschaft und die Entwicklungszusammenarbeit. In: Greinert, W. D. et al. (eds) *Vierzig Jahre Berufsbildungszusammenarbeit mit Ländern der Dritten Welt*. Baden-Baden: Nomos Verlag, 1997, p. 399-408.
- Lauterbach, U. et al. *Jährliches Gutachten zur Evaluierung der Maßnahmen des Internationalen Fachkräfteaustausches*. Frankfurt am Main: German Institute for International Educational Research, 1995 (mimeo).
- Lauterbach, U. *Vergleichende Berufsbildungsforschung*. Baden-Baden: Nomos Verlag, 2003 (Bildung und Arbeitswelt, Vol. 8).

- Lauterbach, U.; Göbel, K.; Lanzendorf, U. *Finanzierung der beruflichen Bildung in Europa, Status Quo und Herausforderungen für die Zukunft. Vergleich unterschiedlicher Finanzierungssysteme in Europa*. European workshop, 16-18 June 1997 in Munich. Cologne: Carl Duisberg Society, 1997.
- Nijsten, M. Mobilität: Qualität von Ausbildungssystemen erhöhen und Austauschmöglichkeiten verstärken. Praxisnahe Beratung und umfassendes Angebot zum Erfahrungsaustausch. *Bildung für Europa*, 2002, Vol. 3, p. 9.
- Owen, J. M.; Rogers, P. J. *Program evaluation. Forms and approaches*. London: Sage, 1999.
- Posch, P.; Altrichter, H. Möglichkeiten und Grenzen der Qualitätsevaluation und Qualitätsentwicklung im Schulwesen. Innsbruck: Studien Verlag, 1997.
- Rossi, P. H.; Freeman, H. E.; Hofmann, G. *Programme-Evaluation. Einführung in die Methoden angewandter Sozialforschung*. Stuttgart: Enke, 1988.
- Rost, J. Allgemeine Standards für die Evaluationsforschung. In: Hager, W.; Patry, J.-L.; Brezing, H. (eds) *Evaluation psychologischer Interventionsmaßnahmen. Standards und Kriterien: Ein Handbuch*. Bern: Hans Huber, 2000, pp. 129-140.
- Sanders, J. R. Cluster evaluation. In: Chelimsky, E.; Shadish, W. (eds) *Evaluation for the 21st century. A handbook*. London: Sage, 1997, p. 396-404.
- Schiffler, A.; Hübner, S. Allgemeine Standards für die Evaluationspraxis. In: Hager, W.; Patry, J.-L.; Brezing, H. (eds) *Evaluation psychologischer Interventionsmaßnahmen. Standards und Kriterien: Ein Handbuch*. Bern: Hans Huber, 2000, pp. 141-151.
- Schultz Pinstруп, A.; Winther, M. (eds) *How to ensure the quality in foreign work placements. A handbook of best practice*. (Supported by the European Commission under the Leonardo da Vinci programme.) Cirius, Denmark, 2002.
- Sedlmeier, P. Jenseits des Signifikanztest-Rituals: Ergänzungen und Alternativen. In: *Methods of Psychological Research*, 1996, Vol. 1, No 4. Available from Internet: <http://www.mpr-online.de> [cited 11.12.2003].
- Stockmann, R. (ed.). *Evaluationsforschung. Grundlagen und ausgewählte Forschungsfelder*. Opladen: Leske and Budrich, 2000a.
- Thomas, A.; Abdallah-Pretceille, M. (eds) *Interkultureller Austausch: deutsche und französische Forschungen zum interkulturellen Lernen*. Baden-Baden: Nomos Verlag, 1995.
- Thonhauser, J.; Patry, J.-L. (eds) *Evaluation im Bildungsbereich*. Innsbruck: Studien Verlag, 1999.
- von Homeyer, E. Mobile Ausbilder schaffen flexible Systeme. Auslandserfahrung für Personalfachkräfte. 20 Jahre Austausch, heute mit 40 Ländern. *Der Ausbilder*, 1998, Vol. 46, No 8, p. 17-20.
- Weinert, F. E. Lehren und Lernen für die Zukunft – Ansprüche an das Lernen in der Schule. *Pädagogische Nachrichten Rheinland-Pfalz*, Vol. 2, 2000, p. 1-16.
- Widmer, T. Qualität der Evaluation – Wenn Wissenschaft zur praktischen Kunst wird. In: Stockmann, R. (ed.) *Evaluationsforschung. Grundlagen und ausgewählte Forschungsfelder*. Opladen: Leske and Budrich, 2000, pp. 77-102.
- Wordelmann, P. *Evaluierung von transnationalen Praktika*. Presentation at Agora 13, Mobility as a didactic tool, 29-30 October 2001 in Thessaloniki.
- Wottawa, H.; Thierau, H. *Lehrbuch Evaluation*. Bern: Hans Huber, 1998.
- Various electronic documents were also provided by Cedefop (all dated 2002).

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-Ilina,
Gerhard Schaumberger

Abstract

Based on an understanding of evaluation as a development process, this study outlines the current debate on evaluation of vocational education and training (VET) in Europe. It defines and explains the criteria used for this exercise and applies them to a number of selected case studies from different EU programmes (NARIC network evaluation under Socrates, assessments of Phare VET programme case studies in Bulgaria, the Czech Republic and Slovakia and evaluation of Leonardo da Vinci I). Findings are then summarised with special emphasis on the instruments, tools, procedures and outputs used across different audiences and themes.

Programmes in central and eastern European countries (CEECs) where the authors have most experience, are used to analyse the evaluation capacities and procedures in place. The analysis considers the flaws and potential of the selected evaluation exercises and their impact on policies. Finally, a future research agenda and policy recommendations identify gaps in knowledge and the challenges for research in evaluating VET.

Table of contents

1. Introduction	165
2. Case studies	168
2.1. Case studies on evaluation of Phare VET programmes in CEECs	168
2.2. Case study 1: Phare programme CZ 93.05	169
2.2.1. The context	169
2.2.2. The programme	169
2.2.3. Programme results and impact	170
2.2.4. Final programme assessment	170
2.3. Case study 2: Phare programme SR 94.03	171
2.3.1. Context of the programme	172
2.3.2. The programme: objectives and methodology	172
2.3.3. Evaluation team	173
2.3.4. Programme results and impact	173
2.3.5. Final programme assessment	174
2.4. Case study 3: Phare programme BG 96.04	174
2.4.1. The context	174
2.4.2. Programme assessment: dissemination and mainstreaming	175
2.4.3. Evaluation team	175
2.4.4. Final programme assessment	176
2.5. Case study 4: transnational analysis of national academic recognition information centres (NARICs)	176
2.5.1. The context	176
2.5.2. The programme	177
2.5.3. Methodology of the evaluation	177
2.5.4. Evaluation team	178
2.5.5. Programme results and impact	178
2.5.6. Final programme assessment	179
2.6. Case study 5: LdV I in the Czech Republic	180
2.6.1. The context	180
2.6.2. The programme	181
2.6.3. Main findings: programme results and impact	183
3. Summary of findings from evaluation reports	187
3.1. Before implementation	187
3.2. During implementation	187
3.3. After the programme	188
3.4. The evaluation process	188
3.5. Methodology of assessment	188
3.6. Involvement of national and international experts	189
3.7. Evaluation culture	189
3.8. Interrupted chain of responsibilities	189
4. Research agenda and recommendations	191
4.1. Research agenda	191
4.2. Recommendations	191
4.3. Evaluation as part of implementation	192

List of abbreviations	194
Annex 1 – Criteria for VET evaluation	195
Annex 2 – Criteria-based analyses	198
References	217

List of figures

Figures

Figure 1: Evaluation cycle	165
Figure 2: Link between projects, programmes and policies	167
Figure 3: Traditional EU policy cycle	190

1. Introduction

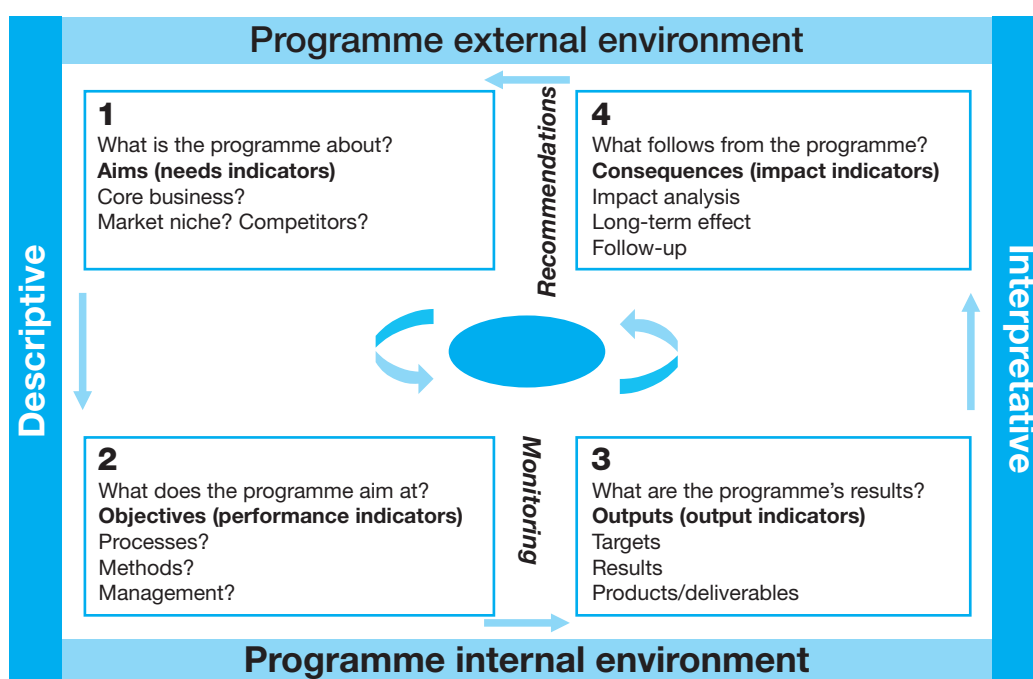
The following definition of evaluation, as developed in the Leonardo da Vinci (LdV) twin projects CERN and EVAL⁽¹⁾, is used throughout this study: 'Evaluation is to determine the significance or worth of something by careful appraisal and study. It is a developmental process that illuminates or enlightens the specific policies, processes and practice of its stakeholders and contributes to collective learning.'

Accordingly evaluation should aim at accountability (measuring stick) and improvement (torch) and cover project aims, objectives, outputs and

impact. It applies criteria-referenced (e.g. defined project objectives), norm-referenced (e.g. against standards) and ipsitive-referenced (e.g. participant's learning curve) description and judgement, internally and externally.

The evaluation cycle for projects and programmes follows a logical sequence (Figure 1). While theoretical understanding of evaluation is closely related to this cyclical improvement mechanism, it is clear that the timing, scope and dimensions of most evaluations do not allow the investigation of all aspects involved.

Figure 1: **Evaluation cycle**



Source: Baumgartl, Navreme knowledge development.

⁽¹⁾ The Capitalisation + Evaluation Research Network (CERN) and the E-VAL electronic environment were established under the Leonardo programme of the EU in 2001. They bring together experts in theory and practice of evaluation from 14 European countries. CERN analyses intention, effectiveness and impact of activity, in order to recommend future action and stimulate collective learning. E-VAL develops a virtual environment to support those involved in evaluation.

The importance of evaluation is increasingly recognised in EU policy-making and is mainly implemented via programmes and projects which, since 1994, have all undergone regular evaluation. Comparison between, capitalisation upon, evaluation of, valorisation from and knowledge development via these programmes and projects are thus key factors in the political and scientific debate in the EU relating to:

- (a) demonstrating value for money in EU programmes;
- (b) making better use of the experience, results and proposals emerging from the significant work undertaken.

Recent discussions about good governance also give evaluation research special policy relevance.

The concept of valorisation entered EU language only recently but has occupied an important place in its agenda. It is now widely accepted in the community of practice of the Directorate General for Education and Culture. The concept is largely bureaucratic in nature: it was introduced and popularised by the European Commission for EU education and training programmes. Although its development included consultation with the expert community, it did not result from a scientific discourse, nor was it seriously validated against research into evaluation. It is defined as 'the process of enhancing or optimising project outcomes through experimentation and exploitation with a view to increasing their value and impact' (European Commission, 2002). However often valorisation is understood as capitalisation on achievements or dissemination beyond the partnership. The concept itself involves monitoring, assessment, dissemination and mainstreaming as integral parts of valorisation, which are at the same time distinct tasks and activities. It is ambiguous: valorisation is often used as a synonym for evaluation or monitoring or dissemination or conversely as a separate activity from those mentioned. In short the term is unclear even where it is most used. To our

knowledge it has not been taken up in European programmes of other Directorate Generals, where monitoring, evaluation, dissemination and mainstreaming of good practices are commonly recognised activities with clearly defined responsibilities. Being a confusing concept, it confused the implementation and accountability behind it. Indeed, are evaluations still necessary as the external and regular activity that provides feedback for further strategic decisions or are they a mere aspect of valorisation? Does valorisation have a formative or summative function? ⁽²⁾.

Many EU funded programmes and agencies ⁽³⁾ have developed substantial monitoring and evaluation procedures for education and training; so have other international organisations (World Bank, ILO, UNDP, Unesco, CoE, etc.). Evaluations are increasingly carried out at national and regional levels (often as a consequence of EU requirements). They differ in detail, but generally follow similar procedures and are realised mostly at the end of programme/project.

Rather than criticising specific evaluation exercises or presuming to judge 'good practice', this study tries to identify 'relevant practice' in evaluation activities and to present the crucial success factors for evaluation.

Moreover, we are interested in shedding light on the different evaluation cultures across Europe and pay particular attention to the intercultural dimension of evaluation exercises, which are often carried out by international or rather transnational cooperation.

Internationally funded programmes and projects are an important tool for policy innovation. While policy development may use various methods to verify relevance of innovative measures and solutions, pilot projects remain the most straightforward and efficient. Equally, without awareness and evaluation of projects and the foreseen impact on policies (often rendered via and fed into policies by an evaluation process), programmes and projects lead nowhere.

⁽²⁾ In 2001, the European Commission established a group of international experts which involved representatives of LdV national agencies, social partners, promoters and national authorities to work on developing the valorisation strategy of LdV II and build on results of the first phase evaluation, undertaken by a group of external experts. The valorisation group was drafted by nomination from the European Commission. The group cooperated with various organisations such as qualification forum, Directorate General Employment. This, however, was often seen as a formal (sometimes even formalistic) consultation process to support arguments of the European Commission. Nevertheless, it contributed to the selection of the best projects and practices with their subsequent publication in the compendium and dissemination at 'Employment week' in 2001 (information from interviews).

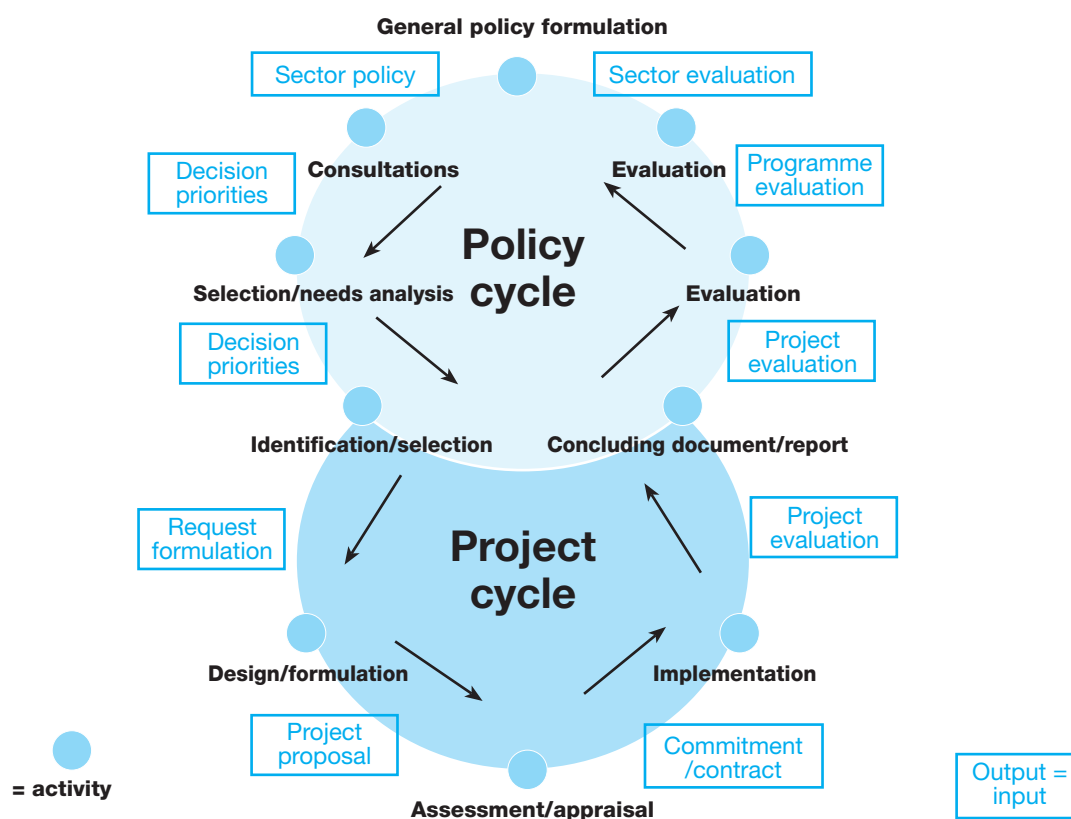
⁽³⁾ E.g. Socrates, Leonardo da Vinci, Phare, Adapt, Youth, fifth framework programme, EU agencies (Cedefop, ETF).

The following scheme addresses the link between projects/programmes and the policies that result from such projects.

We can hypothesise that the policy level benefits from the project implementation only under certain conditions, namely the indisputable interest in project implementation and in evaluation of its design, performance and results throughout the project lifespan. We attempt to verify this statement with the help of case studies of evaluations and assessments performed in different European and international programmes.

The case studies provided in Chapter 2 represent fairly diverse implementation approaches and evaluation methods and procedures. The analysis of the case studies has been performed by means of a commonly identified set of criteria (Annex 1). The main findings stem from the application of each of these research criteria to the case studies and their aggregation into general conclusions (Chapter 3). It is followed by identifying research gaps and recommendations for the European research agenda, policy and practice, in particular for the design of evaluation exercises.

Figure 2: Link between projects, programmes and policies



2. Case studies

Evaluations differ across programmes and countries. Each country has its specific cultural and historical context and we therefore select a range that enables conclusions to be reached. We look at how evaluation modes and cycles differ across various European programmes and in the specific national context and this enables us to identify any strengths and weaknesses of individual procedures.

Evaluation is not separated from policy development and implementation. While an important step in itself, it is only one phase of a larger process. We look therefore at the context, aspects of the legal and political framework, the specific role of the evaluation, eventual links to previous or later stages of the evaluation and the programme itself and finally effects of evaluations.

The evaluation culture evolves with the development of the democratic process and increased accountability for public spending. Evidence of evaluation culture is not apparent among many EU Member States. However the European Commission, by requiring the evaluation of the outcomes and impact of EU funds allocation, has played a role in fostering evaluation activities. CEECs have gained access to several EU-sponsored programmes, which have also undergone various evaluation and assessments.

Under the previous regime, CEECs were not familiar with evaluation in concept or practice. The only stakeholder was the state/party which did not expect any constructive and objective feedback on public action.

We shall now look at our selected cases of programme evaluations as reference points for further analysis. Although we focus on the evaluation documents, we shall refer to other relevant documents and the context of the case study. For each case study agreed common criteria were used to analyse the evaluation reports, the context of the EU programmes under consideration, and their impact (Annex 1). The annexes provide in depth analyses of the first four cases according to these criteria. It should be noted, however, that these criteria do not allow for objective and precise comparison across coun-

tries and programmes. Therefore, we used them rather as a guiding tool for analysis, looking also at additional information and documents whenever necessary.

2.1. Case studies on evaluation of Phare VET programmes in CEECs

Our first set of case studies of evaluating European Community support programmes is the Phare VET reform programme implemented in parallel in 10 CEECs between 1993 and 1998.

Normally, internationally funded programmes, in addition to covering the absence of a fund at the national level, are expected to bring an added value in terms of transnationality. Phare programme are somewhat unusual in that they are mostly designed and implemented at national level. They have been an EU assistance programme to CEECs in their transition to the market economy and democratic society which began at the beginning of the 1990s as part of their preparation for EU membership. Equally, transnationality and European added value were not prime programme objectives, but they shaped implementation significantly. The success of programme actions was often based on the expertise of the EU consortia involved in the process and the European Commission, its delegations in CEECs and various EU agencies. These naturally shaped the design of EU interventions displayed in reform developments in CEECs.

The Phare VET reform programme followed the same principle objectives although the specific design varied in each CEEC involved:

- (a) modernisation of existing and the development of new curricula for training for a range of broadly defined occupations to meet current market requirements.
- (b) support of institutional and policy development in VET and dissemination of pilot school results;

- (c) staff development for a wide range of reform participants;
- (d) upgrading of pilot school equipment.

The programme was piloted in a limited number of selected vocational schools and was intended to be further disseminated to other schools and to policy level.

At the end of the programme an external assessment was carried out in each country by international consultants. Their reports are considered in the following three cases. The reports are technical assessments of processes and outcomes rather than formative evaluation analyses. They follow Phare assessment practices with a formal approach of assessment against the stated objectives and are therefore summative in nature. However, they do not verify the programme impact and effects of certain reform features: these were an integral part of the programme and so the results of the dissemination of the pilot school experience had to be assessed separately.

The following is a short overview of the Phare VET reform programme in three countries, two of which, the Czech Republic and Slovakia, are fairly similar, considering the split of the former Czech and Slovak Federal Republic. The third example, Bulgaria has a contrasting system of education and training and a distinct cultural context.

2.2. Case study 1: Phare programme CZ 93.05

The collapse of the previous regime and of the planned economy system has placed new demands on the skills of the labour force. VET in the Czech Republic has always enjoyed high participation and esteem. However the VET curricula, qualifications and teaching methods and tools were outdated and could not meet the demands of the new economy. Since 1991 programmes such as Labour Market Restructuring, Renewal of Education System and others have been implemented with the assistance of Phare funds.

2.2.1. The context

A 'strategic review of VET' prepared under Labour Market Restructuring suggested recom-

mendations for developing VET in the Czech and Slovak Republics (Birks Sinclair, 1993). Recommendations for the Czech Republic were made to the Ministry of Education, Youth and Sports (MoE). Initially planned funds had to be divided between the two countries after the split of the federation.

Few other national strategy documents were in place but those available were taken into account in drafting the review. The beginning of the 1990s was marked by a strategy and policy framework vacuum and the review was therefore highly valued.

The VET programme was designed to address problems identified in the strategic review. This was not an official policy document and the pressure for change was mainly accession-driven, initiated by the European Commission. It was nevertheless widely supported by actions at a micro level. The period of reform was characterised by bottom-up initiatives, whereas the policy level clearly lagged behind.

The MoE was the principle stakeholder of VET reform. However, it suffered from frequent changes of ministers and key ministry staff. There was no inter-ministerial cooperation, nor was there involvement of social partners.

2.2.2. The programme

The programme was implemented between 1994 and 1998. The Financial memorandum stated the wider objective of the VET programme component: support to development, preparation and restructuring of human resources in and for businesses and public bodies, as required by the economic and social reform programme. The immediate objective was defined as the contribution to long-term reform of VET, as a key element of education, and support of economic reform. The reform programme had objectives defined for each of the seven VET subcomponents, including curricula and pilot school development, staff training, development of up-to-date learning materials, provision of equipment in pilot schools and evaluation of programme results and lessons learned.

Programme implementation was characterised by changes in the reallocation of responsibilities and tasks. The division of responsibilities between the European Training Foundation (ETF), the implementing agency, the European Commission and its delegation was not entirely trans-

parent and resulted in no action in several cases (e.g. mainstreaming the result to the policy level). Others involved were: MoE, Centre for foreign assistance, a national aid coordinator, steering committee with representation of different ministries, advisory task force, a representative body of policy-makers and social partners, international consultants, national experts and in particular the Research Institute of Technical and Vocational Education, which was charged with several tasks and project components. The important involvement of third parties (Ministry of Labour and Social Affairs, social partners and stakeholders for adult training, tertiary education and other educational subsectors) was not fully achieved, especially with regard to implementation of programme results on a larger scale.

2.2.3. Programme results and impact

The investment made in the 19 pilot schools had a significant qualitative impact on their development, however, exposing them to new pedagogical and management practices and teaching methods. It promoted a desire for improvement and efficiency and schools benefited from a study visit abroad and the exchange of knowledge. This encouraged them to take initiatives of their own. Many of the pilot schools proved to be successful project generators and partners in later EU-funded programmes (e.g. LdV, case study 5). A weak point, identified by the evaluation, was that good practices and programme results were not transferred to other, non-pilot schools.

Despite bearing the cost of the programme and its implementation, the MoE did not take up the main programme results and its involvement throughout remained minimal. The evaluation identified this as a major failing not only of the programme implementation but also of its design. The objective of the contribution to the reform of VET did not correspond to the strategy for implementation with its focus on pilot schools and lack of systemic activities. Although the managing unit gave much attention to dissemination activities and public discussion, the lack of receptiveness, interest and acknowledgement of results by the MoE prevented mainstreaming the outputs to national level. In addition the programme arrangement did not involve third parties from the policy level, e.g. the Ministry of Labour, resulting in the lack of linkage to labour-market restructuring and adult training. However recent devel-

opments mark a greater strategy-oriented approach in policy-making with long-term planning and clear targets. The programme results certainly contributed to this and although the impact was delayed, the current strategic policy approach would not have been possible without them. Nevertheless, the legislative framework did not provide for further implementation of the programme results. The new bill of the School Act took into account some of the programme results and approval was sought from Parliament on several occasions. However, due to the lack of political consensus each time, it was decided to hold the document for further elaboration.

The programme's objective of contributing to the restructuring of human resources was not fulfilled, mainly because of the subsectoral approach of the programme, and adult training providers did not participate in the project. VET was tackled in isolation from wider lifelong learning objectives without a specific link to tertiary education, adult training and labour-market restructuring. This was not helpful in meeting the needs of the target groups.

2.2.4. Final programme assessment

The programme assessment sums up the milestones in dissemination activities, including eight thematic reports with analyses of the experience of pilot school activities, three synthetic reports, a draft policy paper (Green Paper), all constituting a major publication *From pilot schools to reform strategy* and a discussion paper *Education for prosperity – Towards the learning society* and others. Although the activities significantly contributed to public awareness, they did not result in improving policy-level receptiveness and results mainstreaming. The programme, however, brought about a new culture of interactive dialogue with a wide range of VET reform participants including social partners and different ministries.

The evaluation (OMAS, 1999) assessed the programme against the initial objectives of the subcomponents as stated in the Financial memorandum. The methodology of the assessment, although not clearly stated in the report, was a fairly standard Phare procedure and responded to the requirements of the *Terms and reference*. The assessors primarily studied the Financial memorandum, the Strategic plan, log frames, the initial strategic report (*The strategic review of VET*,

Prague 1993), work programmes and reports, conducted interviews with project managers at the Project management unit (PMU – National training fund) and on the project sites and with stakeholders of VET reform. The management unit also performed a self-assessment.

The assessment was somewhat formal, corresponding to standard Phare evaluation guidelines. It did not measure the impact, nor did it perform any follow up study at the level of the target group. Although extensive interviewing took place, the evaluation team did not explore principal drawbacks discovered in accepting the content-related outcomes of the programme. There is little evidence of understanding the intrinsic worth of the programme and no verification of its adequacy to the needs in the context of development and strategy. The feedback from such evaluation was, therefore, limited, despite the awareness of the major problems of the programme setting and implementation.

However, the assessment takes into account the risks and assumptions, with regard to the initial setting, identified in the Financial memorandum, which are mostly management and funding driven. The wider policy-related obstacles and risks were not taken into account in the prior identification. The assessment indicates that the absence of a strategy document for the MoE, prepared prior to the programme implementation, adversely affected implementation. In fact even its need in such a policy document was not acknowledged by the Ministry prior to the programme starting.

The lack of providing indicators in the initial documents introduced the methodological problem; the evaluation report often refers to the lack of measurement of programme impact and lack of relevant statistics. While an important issue, this was not foreseen in the programme. It is, however, unclear why no serious attempt in the impact measurement was not foreseen in the evaluation action itself.

There are few links to the strategic review but the log frames were used more actively. This lack of reference can be explained by the formalistic approach of Phare evaluations, where evaluation is targeted to a particular programme, irrespective of its direct linkages to documents and/or activities under preceding or subsequent programmes.

A major drawback in the design of standard Phare programmes is the absence of an *ex-ante*

evaluation, which can thoroughly verify the identified programme objectives, targets and indicators of achievement against priorities and needs based on a comprehensive socioeconomic analysis.

Generally, Phare evaluations were considered by programme participants and partners to be more a threat than a consulting tool. Phare arrangements for evaluations assumed contracting international consortia. And, though these involved some local experts, Phare did not contribute significantly to the development of the evaluation culture in the country, nor enhance expertise in evaluations. Evaluation arrangements were settled within the bureaucratic procedures of the Phare programme, instead of being the responsibility of managing structures and policy-makers. The latter therefore perceived evaluations as imposed externally and, naturally, could not see them as a constructive tool for policy planning and adjustment.

2.3. Case study 2: Phare programme SR 94.03

Slovakia, as a former part of the Czech and Slovak Federal Republic, had been among the first countries to conclude association agreements with the EU. However, at that time the Slovak Republic was not included amongst the first wave of five countries to start negotiations for membership. This was partly due to strained political relations between the EU and the previous Slovak governments, especially during the period from 1994 to 1998.

The Phare programme SR 94.03 was one of the first implemented by the ETF on behalf of the European Commission. It is based on a Financial memorandum signed by the Slovak Ministry of Foreign Affairs and Phare on behalf of the European Communities on 10 November 1994.

The programme intended to complete the new curricula for grades 2-4 of vocational schools, provide teacher training for the new methodologies and provide new technological teaching and learning equipment. EU Partnerships should enable schools to continue with future reform efforts. A strong emphasis on the dissemination of results was to prepare for the mainstreaming of the pilot curricula across the national system.

2.3.1. Context of the programme

The Financial memorandum had been signed by a government that went out of office a few weeks later. The government in power during the implementation phase of the programme appears not to support its general objectives and, in 1996, the Minister for Education was involved in a disagreement over a different Phare programme which overshadowed all Phare procedures.

Nevertheless, the Phare VET programme SR 94.03 gave a necessary, timely and unique contribution to the re-positioning and upgrading of VET in the Slovak Republic. It sustained badly needed reform efforts during a period when educational reform was largely neglected. This success was mostly due to the enthusiasm and effort of committed teachers and school managers. Phare initiatives such as this VET programme have been instrumental in enabling Slovak experts and educators to keep in touch with international developments and in relaunching the political integration of Slovakia with the EU and western Europe.

A key element in the terms of reference for long-term assistance, was the preparation of a national strategy. However such a strategy was never developed as there was no direct involvement of main policy-makers in the programme. Social partners have been reluctant to participate in a project which the MoE itself did not fully recognise. This was apparent at the International conference in Mojmirovce in June 1998 which was meant to function as an important means of dissemination but was merely an information event. The logic for the second seminar on dissemination of results in Povazska Bystrica in August 1998 was not clear either. National decision-makers did not become aware of the critical situation for the continuation of VET reform at that time and consequently no measures were contemplated between the summer of 1998 and February 1999. At the Programme review conference on 5 and 6 February 1999 the Minister reaffirmed his commitment to VET reform and became aware that at some schools the experimental classes and curricula were already discontinued. He immediately undertook to rectify this situation.

2.3.2. The programme: objectives and methodology

The following objectives were specified:

- (a) to provide the Slovak Government and the ETF with an independent assessment of the achieve-

ments of the VET programme, of the relevance and effectiveness of its strategy and activities for the economic transformation and, more specifically, for developing VET with regard to the requirements of a market-driven economy;

- (b) to provide an assessment of the implementation of the programme with respect to its management, its effectiveness and direction with reference to its stated objectives.

The methodology can be criticised in that the lack of monitoring and evaluation instruments has impeded self-assessment and reengineering of the programme content and that the evaluation task was hampered by the absence of planning instruments, a clear and logical framework (matrix), performance indicators or achievement targets. However, the evaluation programme itself tried to balance the lack of information with extensive information collection, visits and by a series of instruments of assessment.

The evaluation was based on the analysis of the following source materials:

- (a) the central monitoring documents of the Phare programme;
- (b) the mid-term evaluation of 1997;
- (c) final reports of PMU and the Long-term Technical Assistance (LTA);
- (d) reports produced within the programme;
- (e) analysis of the programme documentation at the State Vocational Education Institute (ŠIOV);
- (f) visits to pilot schools by the evaluation team;
- (g) workshop reports of the Programme review conference held on 5 and 6 February 1999;
- (h) analysis of questionnaires for pilot school members and non-pilot school teachers;
- (i) inputs by independent Slovak educational experts.

Little or no information was obtained as to minutes or documentation of decisions taken of regular meetings of the steering committee, the advisory task force, the national experts responsible for work of the national working groups, the five national working groups, nor of the meetings with the headmasters of the pilot schools or meetings of members/coordinators of the national experts group and/or national working groups with the respective school coordinators. There are two final reports, one prepared by the LTA, the other by experts contracted by the PMU. Neither was discussed and approved by the programme steering committee.

Great importance was attached to the evaluation proceeding in an interactive way. A significant role was assigned to the available central and eastern European and Slovak experts, with the object of stimulating an intensive national debate on the Phare VET programme and its value for future VET policy-making

2.3.3. Evaluation team

The contract was awarded to a consortium under the lead of Mecca Consulting from Vienna and started on 20 December 1998. The composition of the evaluation team was international, interdisciplinary and wide-ranging. It included experts from the Irish organisation Farrell, Wegimont and Associates, who had already carried out the mid-term evaluation of curricula of this Phare programme in 1997, an experienced evaluator from the Netherlands, an Austrian educational researcher specialising in central and eastern Europe, a VET reform analyst and a PMU Deputy Director of the VET Phare programme in Bulgaria. Slovak members of the evaluation team came from the Vocational Information Resource Centre in the Research Institute of Labour, Social Affairs and Family (VÚPSVR) and the Sociological Research Institute 'Focus'.

Further Slovak expertise came from those involved in the project, i.e. representatives of the MoE, the PMU, VET teachers and administrators from pilot schools, specialists from ŠIOV and independent national external experts who could provide an objective view on the project, notably from the Slovak Academy of Science, from the VÚPSVR, the National Institute for Education (ŠPU) and the Phare coordinator from the Slovak Government office. The Slovak national observatory on VET and labour market helped with management issues and logistics in Slovakia and was responsible for the organisation of the Programme review conference that took place on 5 and 6 February 1999.

2.3.4. Programme results and impact

The VET programme has been designed and implemented on the basis of two assumptions:

- (a) the programme would be successfully implemented in 20 schools;
- (b) through dissemination, the new curricula, modules, teacher training methods and methodological expertise, could begin to be

mainstreamed throughout the Slovak vocational system.

However, it was not clear from the programme design how this mainstreaming was meant to happen. In addition the political climate during the programme was such that there was little attention and interest from the MoE. The impact of the programme activities on key stakeholders was further weakened by the frequent changes of participants and institutions involved. The attitude of the MoE during and after the programme is reflected in the poor integration of programme and system:

- (a) the excellent curricula developed were only accredited for pilot classes and no provision existed for their transfer to other classes or schools;
- (b) the teacher training achievements had not translated into a national practice;
- (c) the EU partnerships and infrastructure procurement could not, for purely financial reasons, be continued;
- (d) dissemination activities, as far as they were carried out during the programme, had not led to a permanent means of circulation and exchange of information.

Two representative findings of the evaluation epitomise this:

- (a) some of the pilot schools (e.g. pilot school 19, the Electro-technical school in Bratislava) had already ceased first grade classes under the now already unofficial pilot school curricula;
- (b) the crucial issue of the diploma for graduates of modular curricula remained unresolved.

The government however, expressed a stronger interest in the valorisation of results, e.g.:

- (a) there were indications that the MoE under the new Minister would carry out a comprehensive assessment of the programme's results and the evaluators urged this to take place as soon as the first alumni finish their studies in June 1999;
- (b) there was a commitment to amend the current diploma award to correspond to the pre-existing diploma (possibly with a reference to the broader and additional education and training the students received).

Under these circumstances, sustainability of the programme was far from assured. If no additional donor support was forthcoming, further use

of the products would depend exclusively on the decisions – and actions – of the MoE. Public statements by the Minister are supportive but the prospect of mainstreaming is still bleak.

The failure to provide the necessary legal framework for the horizontal mode of dissemination through pilot schools results in the fully equipped pilot schools not fulfilling their role. Consequently this potentially powerful tool of dissemination is ineffective in furthering the aims of VET reform. Here ŠIOV intervened with a traditional top-down model of dissemination through publications and conferences. Since the application of educational products of the other programme components were restricted to the pilot schools, this exercise was largely futile. The development of a national VET reform strategy was touched upon by two leading experts during the final conference, but its realisation remained a task for the future.

2.3.5. Final programme assessment

In summary, no systematic monitoring occurred, little transparent documentation was available and there was no procedural or political control. Initiatives took place at the inclination of individuals.

The lack of monitoring and means of evaluation has impeded self-assessment and the reengineering of programme content and the evaluation task was hampered by the absence of a planning instrument, a clear and logical framework (matrix), performance indicators or achievement targets.

In the terms of reference for continuing long-term assistance the preparation of a national strategy was a key element of the programme. For several reasons such a national strategy was never developed. There was no direct involvement of main policy-makers in the Phare VET reform programme.

2.4. Case study 3: Phare programme BG 96.04

The programme was based on a Financial memorandum signed by the then Bulgarian Minister for Education, Science and Technology and Phare on behalf of the European Communities, on 8 March 1996. Programme BG 95.06 was a direct

continuation of the VET component of the previous Phare programme BG 93.14-02 (post-secondary vocational training project). This earlier programme had produced a general strategy for upgrading VET and had developed relevant practice and expertise (including standards compatible with the EU), covering five professional branches.

The need for reform in Bulgarian VET had been often misunderstood as an attempt to underestimate the achievements and misinterpret the difficulties of the Bulgarian system during the transitional phase to a market economy. The title selected for the Phare intervention in vocational education correctly places the emphasis on the need to upgrade, rather than reform, substitute or reinvent an education, training or science system.

2.4.1. The context

The programme's seven components were intended 'to contribute to economic and social reform in Bulgaria through the development of education, vocational training, science and technology and to ensure the availability of national human resources for economic and social restructuring'. Nevertheless, the combination of subprogrammes appears to have arisen from a confused position of the then government and thus provided limited potential for synergy: Upgrading VET, Teacher Career Paths, Foreign Language Training, Financial Management for School Education, Development of Science and Technology, School Drop-outs in General Education, a Higher Education Evaluation and Accreditation Agency, a Science Park feasibility study and the PMU.

In comparison with other Phare VET activities in central and eastern Europe, the Bulgarian VETERST programme had a positive reputation. The programme has supported continuing reform in a difficult and volatile environment. The government's ownership of reform guarantees that its successes will be sustained. The results of the programme have started the upgrading of Bulgarian education, especially to meet the requirements of the *acquis communautaire* to facilitate accession of Bulgaria to the EU.

The evaluation did recognise political and cultural circumstances; for example, three Bulgarian evaluators were included along with the three EU experts.

2.4.2. Programme assessment: dissemination and mainstreaming

The evaluation was based on the analysis of the following source materials:

- (a) the central monitoring documents of the Phare programme;
- (b) the intermediate assessment of OMAS (as part of a general assessment of all Phare activities);
- (c) final reports of PMU and Consultants for all subprogrammes;
- (d) reports produced within the programme;
- (e) analysis of the programme documentation at the PMU and pilot schools;
- (f) interviews at and visits to the Ministries of Education and Labour, pilot schools, regional authorities and institutions set up during the programme by the evaluation team;
- (g) the work groups reports of the programme valorisation meeting held on 14 February 2000;
- (h) analysis of questionnaires for programme recipients;
- (i) inputs by independent Bulgarian educational experts.

Great importance was attached to the evaluation proceeding interactively. A significant role was assigned to the available central and eastern European and Bulgarian experts, with the object of stimulating intensive national debate on the Phare VET programme and its value for future VET policy-making.

Dissemination in the programme was carried out through information leaflets and brochures like *Outlook*, published by the PMU and information brochures published by the project offices. However, the failure to provide the necessary legal framework for the horizontal mode of dissemination through pilot schools created a situation where the fully equipped pilot schools were not allowed to fulfil this role in the best way. The project management stepped in with a traditional top-down model of dissemination through publications and conferences and information packages for non-pilot schools. Because the application of educational products of the other programme components was restricted to the pilot schools, this exercise remained suboptimal.

It is impossible to give a general judgement on the dissemination activities for the VETERST programme because of its subprogrammes and their diverging results, successes and difficulties.

The emphasis on pilot studies and methods, administrative procedures and financial matters reduced dissemination and the impact of continued activities to a marginal subtask. At the end of the programme, the difficult economic and financial environment any follow-up activities had to face adversely affected their potential.

Although the MoE should have accepted mainstream implementation of programme results by formal declaration, no actions were undertaken in this respect. This was openly criticised in the evaluation. However recent developments have marked a more strategy-oriented approach in policy-making with long-term planning and clear targets, the programme results certainly being a factor in this. Although the impact was delayed, the current strategic policy approach would not have been possible without the programme results. Legislation did not provide for further implementation of the programme results at the end of the programme but measures were introduced later. Planning strategy had also been driven by the programming of pre-accession activities and structural funding work.

2.4.3. Evaluation team

The contract was awarded to the Swedish consulting firm AmuGruppen from Stockholm and started on 20 December 1999. The evaluation team was international, inter-disciplinary and wide-ranging. Bulgarian members of the team came from the New Bulgarian University, from the Economic University and the National Institute of Education. The team also included an Austrian evaluator with specific VET experience, an experienced VET university teacher and an evaluator from the UK, a VET reform analyst from Sweden and a British NGO Director resident in Bulgaria while implementing other EU programmes there.

Further Bulgarian expertise came from those involved in the project, i.e. representatives of the Bulgarian Ministry of Education and Sports, the PMU, VET teachers and administrators from pilot schools, specialists from research institutes and independent national external experts who could provide an objective view on the project, notably from the Bulgarian Academy of Science, from the National Employment Institute and the National Institute of Education. The Bulgarian national observatory on VET and labour market (which became part of the Human resources development centre) helped with organisational issues

and logistics in Bulgaria and was responsible for the organisation of the programme valorisation meeting which took place on 14 February 2000 in Sofia.

2.4.4. Final programme assessment

Where programmes are running in parallel in different countries or regions, coordinated support and collaboration between programme coordinators and managers, through joint meetings and networking is invaluable. Similarly, institution building in the competent ministries needs special attention.

Project management capacity needs to comply with EU procedures but the need for administrators to exchange ideas is also to be recognised. The new VET modules required in Bulgaria are similar to those already in use in other countries. These could be adapted and used as a starting point for fine-tuning modules, courses and curricula on Europe-wide key competences.

Knowledge sharing and support needs to continue to be planned for and resourced after the project is finished. Legal requirements, development, dissemination and the increasing experience and abilities of project partners (e.g. school level) need to be recognised and be part of project plans, especially in the case of teachers implementing modular curricula for application in other schools.

It is also important to ensure the cooperation of regional inspectorates and the Institute for teachers' qualifications by including this aspect in follow up plans.

The re-shaping of social partner organisations since the start of transition has generated considerable activity. Education, especially continuous training, is a field where social partners in the EU are assuming a greater role than their original remit. VET could become, therefore, a major interest for social partners as a field where they have influence at all levels. This might raise their status here and in other EU countries help in their development in central and eastern Europe.

Large scale mainstreaming of the pilot results to the national system requires a proper framework:

- (a) consensus on goals and strategies;
- (b) agreement between the main participants;
- (c) adequate funding for implementation;
- (d) clear objectives and a cost benefit analysis for Bulgarian VET.

Sustainability should be achieved by building upon the effect of the programme on the recipients increased expertise and abilities to develop and share with others programme results.

2.5. Case study 4: transnational analysis of national academic recognition information centres (NARICs)

2.5.1. The context

The first indications of the commitment of the European Institutions to mutual academic recognition are to be found in the Treaty of Rome in 1957 ⁽⁴⁾, but it was not until 24 years later, in 1981, that the concept of a network of national information centres emerged and a further two years before it was implemented. The evolution of the NARIC network is fully described from different perspectives in the evaluation.

The main aim of NARIC is:

- (a) 'to remove anomalies, misunderstandings and injustices resulting from the widespread lack of authoritative information on equivalence matters among the Member States;
- (b) to obviate any confusion cases of individual hardship, resulting from the lack of any Community-wide system or network of agreements on the academic equivalence and comparability of the very wide range of diplomas, certificates and qualifications [...];
- (c) to enable higher education institutes to be better informed of the new developments of new types of degree courses in other Member States [...];
- (d) to eliminate a major obstacle currently impeding student mobility in the Community' ⁽⁵⁾.

This shows that the major focus is on the development of NARICs within the EU. In the evalua-

⁽⁴⁾ Treaty of Rome, Articles 49, 57.1 and 128.

⁽⁵⁾ Com(81) 186 final of 29.04.1981 on academic recognition of diplomas and of periods of study.

tion, the NARIC network was also compared to the European network of national information centres on academic recognition and mobility (ENIC), which was established in 1994 by merging the existing, respective networks of the Council of Europe (NEIC) and of Unesco-CEPES ⁽⁶⁾ (NIB).

2.5.2. The programme

The early days of the Erasmus programme, under which recognition centres are financed, can be regarded as the basis for the NARIC network. The system was established under the Erasmus (1989-96) and tested over six years in a pilot scheme involving 145 higher education institutions from all EU Member States and the European Economic Area (EEA) countries. Since the introduction of the institutional contract in the Socrates/Erasmus programme in 1997/98, all European universities can take part in the European credit transfer system (ECTS). The ECTS supports European-wide mobility as an effective means of promoting curricular transparency and facilitating academic recognition. Transparency is created by providing detailed information on the curricula and their relevance to a degree. The main tools used to make ECTS work and facilitate academic recognition are the information package, the learning agreement and the transcript of records. The use of ECTS in vocational training is currently under discussion ⁽⁷⁾. The new structure is being piloted in a number of projects across the EU, both in higher and vocational education.

The responsibility for identifying the roles and tasks of recognition centres rests with the national authorities (MoE). They therefore differ in nature and scope, reflecting the structural make up of every member country of the network. In the early days of the NARIC network, the Commission attempted to define common roles and tasks for the national centres to assist those national authorities developing their recognition centres. Several stakeholders were part of the evaluation. Representatives of the following organisations were interviewed: Euro-citizen action service, Quality assurance agency for higher education, the national unions of students

in Europe, Erasmus student network, European university association, section of the European trade union confederation (ETUC), European forum for academic orientation and the Union of industrial and employers' confederations of Europe.

2.5.3. Methodology of the evaluation

The key issues for the evaluation of the network were:

- (a) an assessment of the relevance and added value of the NARIC network 'in terms of promoting the European dimension in academic recognition';
- (b) an assessment of the efficiency and effectiveness of the NARIC network 'in achieving its objectives and tasks'.

The evaluation covered five years, i.e. 1997 to 2001 and aimed to identify 'useful lessons and recommendations for the future activities of the Network, including its ability to meet the challenges of the development of transnational education in general and of the Bologna process in particular.' Eleven specific evaluative questions were formulated in the terms of reference.

Members of the NARIC network are those national organisations responsible, at national level, for recognition of foreign degrees (recognition centres). The terms of reference made a distinction between the activities of the NARIC network and the activities developed at national level by these recognition centres. The NARIC network emerged, therefore, as an entity distinct from its members; while the network was subject to evaluation, the members were not, as they were subject to the control of their national authorities. The NARIC network had to be assessed as an entity formed by its members (the recognition centres), which is conceptually distinct and different from them. The evaluation was monitored by a steering committee composed of European Commission, Directorate-General for Education and Culture (DG EAC) staff.

The evaluation was conceived as a desk research study but supported by substantial field work. The following evaluative instruments were used:

⁽⁶⁾ CEPES is the Bucharest-based European higher education centre of Unesco.

⁽⁷⁾ http://europa.eu.int/comm/education/programmes/socrates/ects_en.html [cited 3.12.2003].

- (a) a considerable number of documents were examined in order to understand the history of the network, its mission, its legal basis, its developments and activities, the cooperation with ENIC and their network and to understand and appreciate its on-going activities and main achievements. Particular attention was paid to understanding the difference between the NARIC activities of national nature (not covered by the mandate) and those of networking nature;
- (b) non-structured interviews with staff of DG EAC to understand the nature of the subject in question and to gather data;
- (c) analysis of the files and outcomes of the recognition centre projects financed by DG EAC during the period and their assessment by means of a standard 'project evaluation card';
- (d) preparation of a questionnaire sent to all recognition centres;
- (e) case studies of recognition centres;
- (f) an analysis of communication within the ENIC discussion list;
- (g) a textual analysis of the websites of 25 recognition centres;
- (h) interviews with informed external stakeholders;
- (i) design of scenarios.

2.5.4. Evaluation team

The contract was awarded to the coordinator of the Pragmatic network of individual European consultants. The evaluation team was composed of an international consortium of consultants from Austria, Bulgaria, the Czech Republic, Italy, Norway and the UK.

The Austrian expertise came from a Vienna based group of international evaluators. The Bulgarian evaluators had long-standing experience in European projects. The contribution from Italy was provided by an independent consultant, whose key areas of specialisation are project management, *ex-ante* and *ex-post* programmes evaluation. The consultant from the Czech Republic is a researcher at Charles University in Prague with experience in web-publishing and information management. The Norwegian owns a consulting company, specialising in knowledge-based services and regional development. The UK evaluator is a school inspector and university manager.

Eight informed stakeholders were interviewed to gain their understanding of the NARIC network, the nature of their institutional relations with it, their degree of satisfaction and their expectations. These organisations were regular partners or counterparts of DG EAC.

The terms of reference had been produced by the evaluation steering group at the Commission, with input and representation from the recognition centres. The continuous monitoring of evaluation progress, substantial comment on draft and final reports and close attention of the steering group resulted in worthwhile and valuable evaluation results and recommendations. However, direct contact with the recognition centres was limited and took place mostly in a mediated form. Several recognition centres perceived the evaluation exercise as a threat and were reluctant to provide information. Several other recognition centres, in contrast, took a much more positive view which added to the evaluation output. Some recognition centres asked evaluators for a copy of their own interview and recommendations, going beyond the original evaluation remit and adding value to the process.

However, a comprehensive analysis of the network environment was not in the terms of reference and this limited the evaluation and findings to network internal and narrow issues. The multinational composition of the evaluation team proved to be useful as interviewees' requests to be covered by selected evaluators could be met (e.g. the Bulgarian recognition centre refused to be interviewed by their Bulgarian peer who was then substituted by another EU national). The project shows the importance of consulting external stakeholders during an evaluation. If only working with project members, there is a danger that a wider audience will get a distorted picture of the project, its relevance to other projects or developments and its results in terms of dissemination and sustainability.

2.5.5. Programme results and impact

The study revealed that most recognition centres agreed that the Bologna process has significant impact in terms of commitment and workload. With the exception of one NARIC, all respondents agreed that the Bologna process has increased the need for frequent network interactions (this involved both the NARIC and the ENIC dimension) and that the Bologna process is having a major impact in terms of workload.

Impact in terms of commitment was uneven and directly proportional to the part played by the organisation managing the NARIC. Reported were mainly political activities at international and national level and dissemination actions (seminars, workshops). The need for finding common platforms with the European network for quality assurance in higher education (ENQA) had been underlined. Most recognition centres reported considerably increased commitment with only one NARIC recording none.

Finally, apart from one exception, there is a unanimous consensus that the Bologna process has been a catalyst for new opportunities for recognition centres. Many recognition centres refer to increased international visibility of both ENIC and the NARIC network. Individual contributions refer to the following as new opportunities: the prospect of contributing to national policy-making in education; assessment of non-regulated qualifications; development of contacts with labour-market partners to improve information and training on recognition issues; role of recognition in lifelong learning including assessment of informal and non-formal learning; further opportunities to promote the Diploma supplement; increased differentiation of evaluation scale; focus on the cooperation between recognition and quality assurance; need to develop new attitudes to recognition. Some respondents stressed that national authorities should be made aware of the consequences of the Bologna process in terms of workload for the organisations managing recognition centres; urgent countermeasures need to be taken in terms of human resources. 90 % of respondents consider that the follow-up of the Bologna process will have a definite impact on the NARIC network activities in next three years.

2.5.6. Final programme assessment

The NARIC network is available to those national bodies responsible for implementing policies about recognition and transparency of qualifications (the recognition centres). It provides an invaluable forum for the exchange of ideas, expe-

rience and information between recognition centres, at national level.

While the control of single recognition centres is the responsibility of their national administrations, the evaluators considered that the NARIC network had an alarmingly low level of formal, institutional structuring. Its mandate was contained in few words of the European Commission Decision establishing Socrates II ⁽⁸⁾; the role and powers of the European Commission were not enshrined in any formal act. The role and the powers of the NARIC advisory board were unclear, incomplete and not formally adopted. As a result, the network was scarcely visible and its lack of a single point of entry or contact was a source of criticism. Relations with the NARIC network were through the Commission, or one of its Members. As a managed consequence, the distinction between the different bodies (recognition centres vs. NARIC network) and their respective reporting lines, was not always clear.

The inability of the NARIC network to respond to recent developments (notably the Bologna process) had been criticised by several recognition centres and stakeholders. Again, the performance of the network was in question, not recognition centres ⁽⁹⁾ due to poorly defined network structure, roles of those involved and rules.

There was a suggestion that the scope of the network should be expanded beyond that assigned by the Socrates II decision. This development met the wishes of most of those involved, but required a formal process leading to the adoption of comprehensive terms of reference for the network, the identification of participants, the allocation of clear roles and responsibilities and the approval of internal rules. Structured long term planning cycles also needed to be adopted and, in the absence of such processes, the development of the network was destined to be a source of further problems.

During the evaluation facts emerged that suggested that several recognition centres were severely understaffed. This could have affected the performances of the NARIC network, which was the responsibility of the European Commis-

⁽⁸⁾ 'In particular, it will collect and disseminate authenticated information which is necessary for the purpose of academic recognition, also bearing in mind synergies with the professional recognition of diplomas.'

⁽⁹⁾ The situation, although not analysed because of the mandate received, was different in the case of several recognition centres, which reportedly had been quick to adapt their structure and some of their working practices to the recent developments.

sion. The NARIC network and the European Commission were advised to take appropriate action. The European Commission had earmarked funds for financing activities related to the recognition of diplomas. Their impact on the overall Socrates budget decreased over time from 0.16 % (1995) to 0.09 % (2001). They were being used to finance activities of different nature:

- (a) projects of national interest;
- (b) projects of bilateral interest;
- (c) projects of network interest.

Projects of networking interest had been relevant overall to the scope of the network and mainly proved to be effective, with a positive impact on the network and the activities of the recognition centres. Projects of national and bilateral interest were designed to support the recognition centres locally (i.e. nationally) and had minimal impact on the network. This helped in strengthening the national dimension (the recognition centres), while the corroboration of the very network dimension was less supported.

The relationship of yearly priorities and projects to be financed must be observed and adequate inter-network collaboration developed.

There was a strong sense of identity among recognition centres, going beyond the NARIC network and covering all countries where ENIC is represented. The recognition centres largely cooperated with each other and were responsive to requests for support from colleagues in other countries. This is a key element for success in redefining the competences of the network. The many differences in roles, national support and staff of the recognition centres have to be taken into account if they are to be able to support and implement the activities proposed by the network in a pro-active way.

2.6. Case study 5: LdV I in the Czech Republic

2.6.1. The context

LdV ⁽¹⁰⁾ is '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' (European Commission, 1994, p. 10). Here is the first inconsistency: being transnational in nature, the programme seeks to keep the status quo in national responsibilities and the diverse systems of education and training but assumes an added value from the promotion of the European-type transnational project.

The Council Decision sets out a common framework of objectives, measures and priorities. The programme priorities and measures were identified prior to the start of the programme and took into account experience with previous community actions (Commett, Eurotechnet, Force, Petra, Lingua). The compatibility with national priorities and strategies was assured by the representation of the Member State in the programme Committee (see further). The common framework of objectives and the list of measures and priorities were defined in an extremely broad manner. While this permitted the accommodation of interests and needs of various target groups and countries with the diverse character of national systems, this approach lacked a clear focus and encouraged supply-driven projects.

The Council Decision encouraged Member States to ensure a balance between the LdV programme and other initiatives in VET. This was also stressed in the allocation of technical tasks, being complementary with structural fund measures, bearing in mind that the latter are intended to match national priorities and strategies for policy development and implementation. The distinctive character of the LdV programme was determined by two basic requirements: transnationality and innovation. The consequences for evaluation will be covered later.

The Council Decision also set out the requirement for evaluations 'at regular intervals on a partnership basis involving the Commission and the Member States', including the results of Community measures that were to 'be evaluated objectively by outside bodies at regular intervals

⁽¹⁰⁾ The case study deals with the first phase of the LdV programme only.

[...]’ (European Commission, 1994, p.14). However the procedure was defined in rather vague terms where a proposal was to be made by the Commission but the Programme Committee was to give its opinion on the matter. Consequently, the evaluation procedure for the programme interim and final results and the programme effectiveness, efficiency, adequacy of measures to the needs of target groups and other evaluation aspects were not clearly defined at the start of the programme.

The Council Decision (1994) opened the LdV programme to the associated CEECs and after the appropriate preparatory work the countries gained full membership in the programme. The Czech Republic has been participating in the programme since 1996 with a full-membership status since 1997.

2.6.2. The programme

Unlike Member States, the CEECs’ entrance to the programme was preceded by an *ex-ante* study aimed at identifying national priorities in VET, the capacity of the systems and arrangements of these countries to participate in the programme, an *ex-ante* evaluation of the potential impact of the programme on policies and practices in these countries and formulation of specific proposals for the participation of these countries in the programme (Handley et al., 1996). Although the study was prepared by a team of international experts, it involved drafting national background reports by local experts and a fairly broad consultation process with national authorities of the six countries involved in the study (Bulgaria, Czech Republic, Hungary, Poland, Romania and Slovakia). The study analysed the major challenges arising from the related socio-economic and legislative situation in the countries concerned and drew up specific recommendations with regard to the identification of country-specific priorities and programme procedures to be implemented in central and eastern Europe in order to maximise the programme impact in these countries.

The study highlights the delicate nature of the transition process. The reform of education and

training had to tackle the major issue of contributing to the overall democratisation of the society and upgrading and adjusting the outdated education programmes to the quickly changing needs of the transforming economy. It also had to face the challenge of preparedness to accession needing to catch up with requirements and strategies at the EU level. The study foresaw, therefore, the likelihood of ‘an impact on internal priority needs as well as those associated with the preparation for participation in the wider European family’ (Handley et al., 1996, p. 4).

The broad nature of the LdV programme allowed the national priorities of CEECs to be accommodated easily, though with certain specificities. For instance, the programme being ‘a transnational laboratory for innovation’ did not assume a linear transfer of know-how from west to east (this in fact was largely done in the framework of the Phare programme). However such a transfer was a natural need, especially at the early stage of the programme. This was not only because of objective underdevelopment but more often a lack of experience in transnational collaboration and project management skills and low self-confidence on the part of central and eastern European institutions and individuals⁽¹¹⁾. The LdV programme expected a European added value, while for the CEECs national level impacts could be more important.

The suggestion of the *ex-ante* study (1996) to take into account the national priorities in the preparation of national calls for proposals and selection of projects was, therefore, in conflict with the LdV programme, which was oriented on transnational action and simply lacked a procedure for giving more weight to national priorities. The study pointed out the need to reinforce the acknowledgement of national needs of candidate countries, bearing in mind that in many CEECs LdV represented a major financial source in VET. It also suggested that scarce financial resources be maximised by selecting the projects with a systemic and/or policy impact. It failed, however, to suggest means of achieving such a requirement in the procedural context of the programme.

(11) This can be confirmed by the fact that with increased experience in the programme participation, some central and eastern European promoters proved to be successful in competing in project selection and in introducing innovative ideas and products in the transeuropean perspective.

The transnational nature of the programme and the lack of a mechanism which recognised national priorities do not explain why CEECs were not properly consulted in the annual identification of priorities in the standard LdV calls. The representation of the opinion of Member States was ensured through the programme Committee, which constituted a body assisting the Commission in the programme delivery, composed of two representatives from each Member State (European Commission, 1994; Article 6). CEECs' representatives have had the right to participate in the work of the Committee since 1998 but with observer status only and on a number of occasions were not given the opportunity to express opinions at all ⁽¹²⁾.

The representatives of the LdV national agency considered that the national priorities at the beginning of the programme were not taken into account because there was limited project capacity at the level of promoters. The National Coordination Unit's strategy of including as many and varied institutions as possible resulted in a great number of institutions, previously inexperienced in international cooperation, gaining experience of international cooperation and knowledge of project management. A project culture has emerged leading to the success of Czech projects in recent LdV calls. Development of project generation and management skills is also important in respect of future structural funds intervention ⁽¹³⁾.

The study pointed out that many Phare projects were pilot in nature and therefore micro-focused and that LdV may develop similar problems, if the appropriate dissemination and effect multiplying activities are not foreseen at an early stage. The major impact of the programme was identified as the new partnership based method of cooperation with Member States institutions (unlike donor-recipient relations in other programmes) and in the forging of new relations with the EU. Wider impacts than those arising from the VET innovations were therefore expected.

Because of the fluid and dynamic nature of developments in the CEECs, the *ex-ante* study

proposed annual reviews and subsequently adapting national priorities, ensuring clear links with national strategies in VET. Achieving the optimal impact on national systems assumed a number of arrangements in monitoring and evaluations. The study suggested evaluating the implementation experience after the completion of preparatory measures ⁽¹⁴⁾ and in the second-year of projects proposed by CEECs and, in fact, assumed continuing evaluation as such. The *ex-ante* analysis also indicated a lack of evaluation capacity in the CEECs and the need to reinforce evaluation skills and work by training the national coordination unit staff. Also needed were appropriate instruments and clearly identified rationales, procedures and beneficiaries for evaluations (Handley et al., 1996; p. 55). This, however, needed a plan from the European Commission with a distinct allocation of responsibilities and evaluation activities at various levels. Such a strategy was lacking and the terms of reference for national coordination units was very vague with regard to evaluations. It was not clear whether any regular evaluation assignment is expected or whether any specific evaluation will be needed on request from the European Commission. In any case, monitoring and evaluation in LdV I was not designed as a comprehensive system with a clear framework, indicators and methodology.

The study results and suggestions were hardly taken into account at later stages of the programme, nor were they consulted in the national evaluation activities. The overall justification of the study is therefore unclear, although main findings appear both adequate and useful. The programme also did not produce a complex *ex-ante* analysis for all countries involved (EU and central and eastern Europe) before the start of the programme and prior to formulation of priorities and measures. The expected impact of the programme therefore was never identified at an early stage as clear quality indicators and quantified goals were not in place. These had an important effect on later evaluations and the programme

⁽¹²⁾ Results of interviews with Czech LdV national agency representatives.

⁽¹³⁾ Results of interviews with Czech LdV national agency representatives.

⁽¹⁴⁾ The country dossiers initiated by ETF covered major elements of VET and LdV priorities of candidate countries contributed to preparatory measures of the programme. However, follow up evaluation of benefits of LdV to the national needs was undertaken and the dossiers remained a self-existent informative document.

attained almost a linear cycle, where the provision of the feedback from evaluations for adjustment and strategic planning was limited.

2.6.3. Main findings: programme results and impact

Article 10(3) of the Council Decision asked countries to produce 'national reports' on implementation of the first phase of the LdV programme. All Member States and some CEECs produced such reports between 1999 and 2000. Not all CEECs were engaged in the activity because of their delayed involvement the programme and therefore finalised projects largely did not exist. For the same reason CEECs were not involved in the full-scale impact assessment and presented an interim report produced either by their national coordination units or by an external evaluator. Clear evaluation guidelines were not provided to the national level nor was it clearly stated whether the exercise should be performed externally or internally.

The Czech national coordination unit decided to perform two tasks. First, it produced a report which focused on the purpose of the programme, its results and influence on VET and employment in the period between 1996 and 1999 with a view to producing recommendations for the next phase of the programme (NCU/NTF, 1999). Second, an external evaluation was produced by three independent experts in 2001, specifically aimed at valorisation of major results and best practices of the programme in the Czech Republic (NA/NTF, 2002). Each of the documents had a different aims which was of worth in itself. While the first was formative in nature, the second presented some aspects of the summative evaluation.

The first report (1999) followed the European Commission outline and provided some contextual analyses of education and training and assessed the impact of the programme in quantitative and qualitative terms. Because projects were still in progress, a proper impact analysis was not possible. The report did not verify the estimated impact of the *ex-ante* analysis nor could it verify the qualitative or quantitative indicators of achievements, as those had not been set out at the earlier stage. It did however arrive at some important conclusions. As assumed in the *ex-ante* analysis, the programme had its major

impact at the micro level. The major benefit of the programme was considered to be in the placement and exchange measures, particularly in respect of the improvement of individual's professional competences and foreign language skills.

Acquaintance with foreign VET, project-design and management skills and transnational cooperation were significant aspects of the programme. The results reported came mostly from regional and local VET. At the same time the programme showed evidence of the rapid increase in the number of Czech institutions involved in the programme during the fast-changing and unstable conditions of the transition period and in the shorter period of the programme compared to Member States. The programme also resulted in a number of valuable innovations, especially new educational products (modules, study materials, educational software, etc.) and contributed to harmonisation of Czech methods in VET with those in the EU Member States. The report assumed that these innovative results will later be adopted not only at the regional and local levels but also by sectors and at national level.

In many cases this was subject to the receptiveness of the national authorities and to legislation. The report did not provide the analysis of major barriers for mainstreaming good practices and innovations. It pointed out the qualitative change in the political context where major strategic documents on education and employment were explained, publicly discussed and accepted at the policy level. However, these documents provided a general vision, a strategic framework, while the legislative side remained stagnant and unreceptive to innovations. For example, many projects focused on development of modular training while the bill of the new School Act, which gives the opportunity to introduce modular training, was rejected by parliament. Mainstreaming of innovations, therefore, is often subject to legislative openness and the ability of systems to take on board the results of experimentation. The analysis of existing systemic and legal barriers appears therefore to be very useful in future.

The second report, an external evaluation study of 2002, clearly presented a methodology developed by the National agency ⁽¹⁵⁾ on the basis of the experience of Member States that

⁽¹⁵⁾ The National coordination unit was transformed into the National agency in 2000.

conducted such evaluations earlier. The evaluation aimed at a twofold objective: assessment of the impact of the first phase of LdV in the three aforementioned areas at the local, regional and national levels; and identification of the best quality products and recommendations for valorisation. This report, along with a number of valorisation seminars, contributed to the valorisation initiative of the European Commission.

Valorisation in the Czech Republic was focused on identifying projects with a high potential for exploiting their results on a large scale and on a wide dissemination of information (NCU/NTF, 2002; p. 56). In addition to the evaluation report, the Czech national agency ran a number of dissemination workshops and a conference and drafted standardised project fiches for projects identified as of high potential impact. The authors of the evaluation report highlighted the problems of carrying out impact assessment when many projects were just being finalised or were only recently completed: a longer-term perspective for the impact evaluation was felt as a need. Therefore, the evaluation report focused on a potential, more than an actual, impact. This corresponded to the proactive valorisation strategy introduced by the European Commission. Nevertheless, the real impact may be useful to evaluate in the *ex-post* evaluation, although such a phase does not seem to be part of the LdV programme sequence.

The evaluation (2002) was voluntary and carried out at the end of LdV I projects. The methodology and the terms of reference for evaluators were drafted by the National agency, as no joint methodology was provided from the European Commission ⁽¹⁶⁾. The evaluation in the Czech Republic was performed by three external experts, each representing an area of LdV: secondary and higher vocational education, tertiary education labour market and employment. While the decision to allocate the task to three independent experts in the three, possibly most, important areas of the programme was methodologically straightforward and an intelligent move, it contained one drawback: such a division prevents a complex look being taken at the programme impact by focusing on the linkage between training and employment or the contri-

bution of innovations to lifelong learning in initial and continuing training.

The evaluation (2002) did return to the identified impact in the *ex-ante* study. However the Czech part of the *ex-ante* study was very weak and only the synthetic analysis could be used as a point of reference and reflections. This however was not required by the European Commission. If the report of 1999 followed the programme results according to the shortened and simplified version of the common framework of the programme objectives, this was not the case of the later evaluation (2002). In fact neither of the two concentrated on programme achievements by priority and measure. This is unsurprising as there were too many objectives, measures and priorities to perform a focused analysis. The evaluation, lacking the overall strategic framework for the programme at national level and therefore lacking the initial point of reference for the programme evaluation, was attempted to be incorporated into the broader strategic framework. It compared the programme objectives with the national priorities identified in two major and relatively recent documents: *National programme of development of education* (the White Book) and the *Human resource development strategy*. A considerable correlation of objectives and priorities was identified, which clearly demonstrated the contribution of LdV programme to national developments. Methodologically this was an instant solution which attempted to compensate for a thorough *ex-ante* analysis of the national strategy and priorities with regard to the LdV programme and the specific needs of target groups in the Czech Republic. The situation is exacerbated by the Czech Republic's lack of strategic policy vision and practically no crucial policy documents. The policy vacuum at the start of the programme was not, therefore, compensated for by a proper needs analysis or any strategic programme planning.

The programme evaluation of 2002 in the Czech Republic did not include the context analysis, nor the analysis of the legal and system framework. Consequently, although many successful projects and useful products were identified, it was not possible to estimate the likelihood of mainstreaming results, despite efforts

⁽¹⁶⁾ Information from the interviews with the National agency representatives.

by the National agency to extend the life of successful projects afterwards. It is well recognised that the best innovations can be lost if not incorporated into legislation and in the Czech Republic many innovative products are at risk due to the lack of legal provision for modular training, competence-based systems of qualification, continuing training of adults and the weak role of social partners and little financing. Unfortunately, neither national evaluation documents of LdV provided a thorough analysis of the barriers to mainstreaming best practices and products and ways of overcoming them.

An important finding of the programme impact was that international cooperation presents a value in itself. While implementation of international projects is complex and challenging and has to take into account legislative, cultural and systemic contexts, it is very rewarding. Meeting these challenges helps build new skills and provides new experiences among those involved. This was important for CEECs, whose experience in generating and implementing international projects was only recent. As with the previous evaluation, the major programme impact was identified at the micro level, i.e. at the regional, local, company or individual level, depending on the project and the specific problem tackled. In the latter case, most projects stemming from identified problems naturally tackled Czech-specific problems. Many projects however, exceeded expectations, with a special European dimension. Many innovative products were developed: elaboration of modules and credits, handbooks, multimedia training materials, methodologies of quality assessment, etc.

An interesting feature of the Czech projects was noticed in the evaluation (NCU/NTF, p. 37): innovation was more common in teaching methods than in the content of education not just in those projects where Czech institutions were partners but also for Czech-generated projects. This was because the Czech institutions, although they generally followed their more experienced western partners' priorities, also had their own. This was also a pragmatic approach, with fewer obstacles expected in mainstreaming teaching methods compared to innovated educational programmes. Many upgraded programmes produced under the Phare programme remained 'pilot' for a long time, lacking the interest of policy-makers and a legisla-

tive mechanism. System settings sometimes present barriers in disseminating good practice on a larger scale. For example, LdV did not produce the results initially expected in linking teaching and research in Czech higher education. As in all CEECs the division of research and teaching in higher education was due to the Academy of Social Sciences being a dominant research organisation and the existence of a number of sectoral institutions with narrow-focused and applied research and development activities. Although the process of transformation saw the reduction of much sectoral research and the role of the Academy diminish, the system overall continued and was major obstacle in the large-scale implementation of research and innovation in universities (idem, p. 38).

The likelihood of cooperation and the effectiveness of projects involving universities and SMEs could be improved by the State by proactive policy measures aimed at supporting research and development in small businesses. The report suggests more support measures are needed in creating requirements for the transfer of knowledge and innovations by developing networks of SMEs with high innovation potential. The report cites as an obstacle: 'The surviving orientation of the State on solving current short-term problems in large industrial corporations or in the banking sector is a serious obstacle in our conditions. Innovative enterprises are not sufficiently supported and they do not stand in the centre of interest of the political elite' (idem, p. 39).

The evaluators concluded that this situation does not support the multiplying effect of innovative projects among SMEs, so the impact is small. The few projects that act on the situation and try to make use of foreign experience are of great importance. Changes in the content of education, quality of practical training and human resource development and increased product competitiveness are important results.

The report indicates that the greatest impact was achieved by projects that attempted to meet the needs of a broader group of institutions, beyond the partnership (idem, p. 66). It can be concluded that inadequate results and limited impact are often 'caused not by unsuccessful realisation of the project, but by a significant difference in conditions in which the project aspires to be successful' (idem, p. 39). It is, there-

fore, important to have in place a thorough analysis of such conditions before commencing the programme and monitoring and regular evaluation focused on those conditions. Further, impact does not have to be about systems, it may be more important at a micro level, i.e. at the level of institutions and individuals. The contribution of LdV to the experience in international cooperation and generation and implementation of European-funded projects should not be underestimated. The impact at individual level goes beyond the acquisition of new skills and is evident in its effect on cultural and professional values.

Evaluation and valorisation documents define the programme too broadly, making focused impact evaluation almost impossible. The European Commission, therefore, suggested in-depth training needs analyses and diagnoses as a useful tool for better targeting (European Commission, 2002). A need for a more demand-driven approach is now being discussed on the programme agenda.

As far national evaluation reports are concerned, they contributed to some extent as reference material to the external evaluation of the programme at EU level, to the final report on the first phase of the programme (European Commission, 2000) and to the valorisation strategy (European Commission, 2002). They were widely used for promoting the programme, but not so much for further strategy adjustment and priority formulation. The latter happened as a

result of the sequence of evaluation and planning. Discussion of the LdV II and the decision to open the programme took place in 1999 and could not take into account the results of the reports. The first call was prepared at the beginning of 2000, the priorities for which were identified for the period of 2000-02. It partially took into account the national reports prepared in 1999 ⁽¹⁷⁾.

Nevertheless, many things needed to be changed and this was confirmed by most evaluation documents. The recommendations were to limit the number of objectives ⁽¹⁸⁾, to simplify procedures and to decentralise implementation of the programme, in particular selection procedures and monitoring, and these were subsequently implemented. The methodological deficiencies in evaluation at national level caused difficulties for the overall programme evaluation. A lack of longitudinal studies, control groups and follow-up studies were characteristic of the programme not only in central and eastern Europe but also in the majority of Member States. There are examples of good practice in evaluation and related research that need to be disseminated to other Member States and the evaluation experiences of other EU-funded programmes may provide inspiration. Nevertheless it has been recognised that LdV needs to adopt a more demand-driven approach in evaluation with clear procedures, guidelines and methods for the subsequent phases of the programme.

⁽¹⁷⁾ Results of interviews with National Agencies.

⁽¹⁸⁾ Formulated on the basis of the working group selected by the European Commission and national authorities, apart from the CEECs representatives. The Programme Committee approved the objectives with CEECs representatives as observers only.

3. Summary of findings from evaluation reports

Using the criteria-based analysis (Annex 1) to structure the findings, we present the summary in chronological order according to the major stages of programme/project evaluation.

3.1. Before implementation

If the programme design has weaknesses, for example, a restricted definition of VET, a bottom-up approach to development at the expense of policy level involvement, the absence of effective dissemination mechanisms of results into practice, this will result in implementation failings. The 'Log-frame' tool ⁽¹⁹⁾, which identifies major opportunities, threats, risks and assumptions, proved very useful. However, in certain cases the method and its structure are too vague to identify major contextual challenges and design an adequate strategy at all levels. The frequent failure to provide indicators in initial documents introduces methodological flaws and is a major problem for measuring progress, monitoring and evaluation. Only a thorough *ex-ante* evaluation with analysis of main socioeconomic trends, labour-market requirements and institutional settings with a consultation process going in parallel, creates the conditions for consensus on what the programme has to achieve and where the reform process should go. If this condition is not fulfilled, national authorities often do not identify with the results of an international project and therefore do not take them on board for further mainstreaming.

The pilot school approach adopted in some Phare projects is innovative but as they were weakly linked to the policy level, they did not lead to the systemic change that was hoped for. A more systemic and policy oriented design of projects is recommended with pilot schools as an integral part of a larger concept at the heart of future projects.

Project implementation is suboptimal when the involvement of the political level (e.g. Ministry) is minimal. Evaluations have identified this as a major drawback not only of programme implementation but also design. Also the NARIC and the LdV cases show clearly that the new institutions need to be as close to the national ministries and regional political decision-makers as possible, in order to have the mainstreaming impact desired. There is also the continued risk of losing the transnational perspective, which can lead to poorly functioning European and international networks.

3.2. During implementation

As seen in the case studies, the complexity of programmes creates difficulties in identifying appropriate and measurable indicators of achievement. It is essential that the programme design contains effective monitoring and continuing evaluation mechanisms in order to fulfil the formative function for success at policy level.

For successful and smooth implementation it is essential that the programme design takes account of cultural factors and institutional, systemic or personal characteristics of those involved. For example the lack of a tradition of constructive inter-ministerial dialogue and systemic provision for social partner's involvement are often added flaws in VET reform projects. Reform efforts were especially successful in those countries where the national, regional and local levels had jointly worked to common targets in a complementary top-down and bottom-up approach. Definition of these targets needs major stakeholder involvement at all political levels, from the macro to the micro level, making the whole reform process more transparent.

Dissemination is crucial in making project outputs available to those not involved (other

⁽¹⁹⁾ The logical framework approach, for example: <http://www.worldbank.org/evaluation>.

stakeholders, wider target group, constituency). Even where dissemination activities and public discussion take place, any lack of receptiveness, interest or acknowledgement of results from the responsible Ministry prevents mainstreaming of programme outputs at national level. Different formats, presentations and mechanisms are essential in reaching stakeholders and the wider audience with information. Systemic dissemination mechanisms incorporated into the initial programme design are helpful for mainstreaming and cascading best practices throughout the project. The difficulty occurs however, in defining best practice and meaningful results prior to final and *ex-post* (impact) evaluation.

3.3. After the programme

The evaluation exercise is less likely to produce relevant results if the distinction between programme and impact evaluation is unclear. Legal measures and the forming of institutions are crucial in achieving sustainability (institution-building).

There should be a strategy to check the extent to which published results and suggestions from project evaluations are implemented in the target systems later. The term 'mainstreaming' of pilot projects means repeating project experience and results in other contexts and making them become policy, backed by appropriate measures.

3.4. The evaluation process

For it to be successful, evaluation needs to be understood as a development process, which aims at capacity and institution building and creates self-evaluation skills among those affected. In many cases evaluation results are only made available to the project partners (partner countries). A process for a more rational dissemination to all Member States should be implemented.

The difference between programme and impact evaluation needs to be stressed again. Impact evaluation requires taking account of the context: political, economic, institutional and social. If not carried out initially in the framework of an *ex-ante* evaluation, at the end of the programme any impact evaluation has no starting point for analysis.

Evaluation budgets are often too small to research project achievement fully (as a rule, 5 % of the budget for small projects and 3 % for large programmes should be made available). This aspect is also linked to one of the more time and cost consuming sides of evaluation, the dissemination process of the results. Dissemination is the key activity in provoking interest in and preparing for mainstreaming to non-pilot schools ('cascading' via demonstration events or information packages).

Dissemination can support the major aim, sustainability. Sustainability means delivering benefits to the target group over an extended period (visibility, networking, policy consensus, user's feedback, funding, official recognition, competent staff, human resources commitment, remaining outputs). We therefore consider that the products of evaluation should go beyond the usual workshop/conference and report and a variety of products to address different target groups and stakeholders would increase their learning from the exercise.

3.5. Methodology of assessment

The method is primarily dependent on the case. Not all Phare and other international programmes can be evaluated using one single method. However, there are some common strengths and weaknesses of methodology.

First, *ex-ante* evaluations may be necessary before the launch of major EU programmes at national level (or feasibility studies enhanced). In certain cases *ex-ante* evaluation may need to go beyond the assessment of existing priority themes against the socioeconomic context and identify priority themes and measures on the basis of the context analysis.

Second, careful description of indicators is necessary. Time spent on project and indicator design is well spent as it will save resources during implementation prevent difficulties of monitoring or incomplete evaluation. The suggested methods of self-evaluation required by the European Commission from the project promoters are still vague.

Third, evaluation needs to be perceived by all stakeholders as a development process: it should aim above all at capacity and institution building and sustainable self-evaluation skills. It is better to reduce the scope of the evaluation and

complete a whole circle of reforms rather than attempt too much at one time and leave crucial topics incomplete (e.g. number of professions selected; number of pilot schools; etc.).

Continuous monitoring of evaluation progress, substantial commenting on draft and final reports and close attention by the steering group increase the validity and accuracy of the evaluation and its recommendations for the stakeholders. The central State and public administration in central and eastern Europe lacks experience in evaluation practices, methods and mechanisms. Overt manipulation of evaluation input forms, reports and programme documents also takes place, resulting in inaccurate evaluation exercises and findings.

3.6. Involvement of national and international experts

The more that national experts are involved in all stages of the programme, the more transfer of knowledge takes place. This way, the evaluation culture introduced by the Commission can be transferred to the national level. National experts should be involved in all components of the programme. They are the major source for future sustainability, enduring knowledge transfer and capacity-building.

However, the multi-national composition of the evaluation team was helpful in allowing specific requirements of interviewees to be met by selected evaluators. The outside perspective of international evaluators needs to be maintained in order to guarantee national comparative evaluation expertise and to promote the European dimension.

3.7. Evaluation culture

Evaluation abilities need time to develop. 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.

The translation of evaluation findings into policy often lacks systemic links. Evaluation findings and reports, especially if critical, may be shelved or not disseminated at all.

One positive finding is that the programme and evaluation practices introduced by the EU in candidate countries and Member States caused a major drive to strengthen evaluation culture. The practices were often adopted by the respective States for evaluation of international and national measures and in candidate countries this is a slow but promising process.

3.8. Interrupted chain of responsibilities

Attention needs to be drawn to one aspect which is considered a flaw: the split of responsibilities throughout the programme cycle via a chain of contracts with 'limited responsibility'. There is no overview of the programme from its inception to final recommendations. Responsibility is shared at all stages necessitating exchanging information, solid monitoring and reporting practices and formats and the assignment of single project directors throughout project duration.

The following identifies all the steps necessary to conceive a Phare traditional programme for a given country.

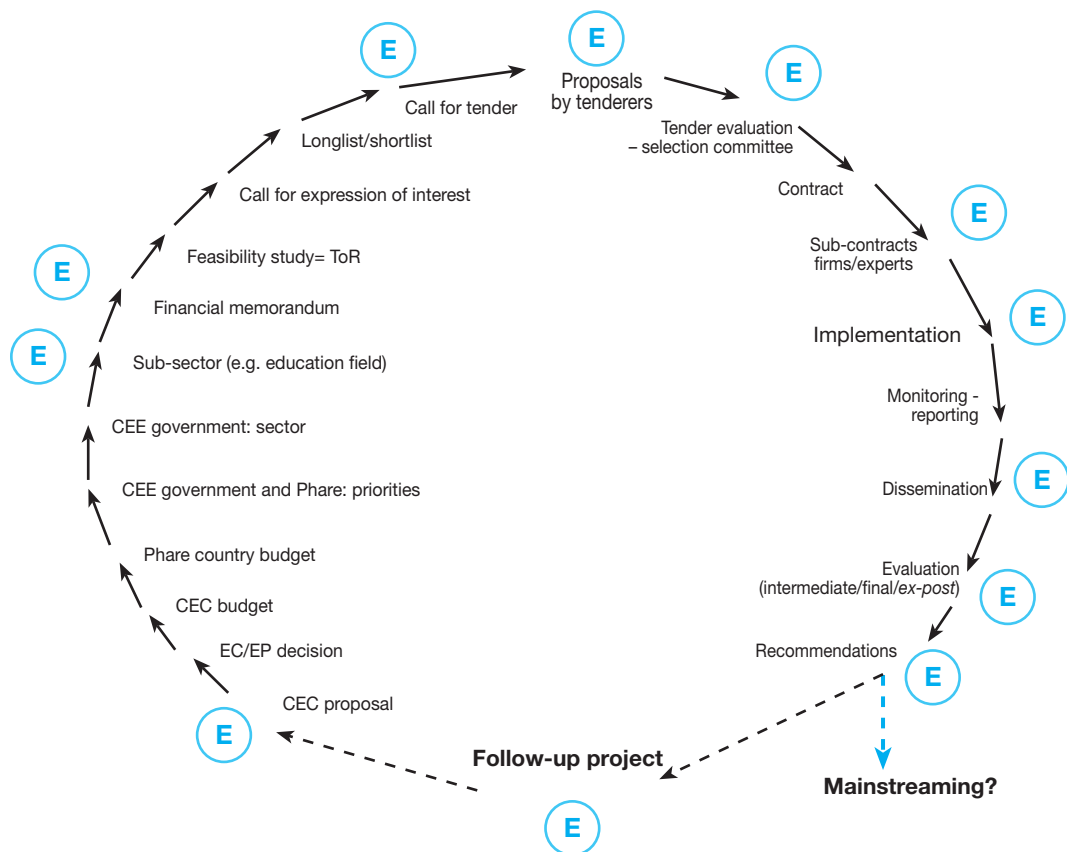
Clearly, the split of responsibilities at each step results in a lack of overview and control of the whole process. 'E' in a circle, in the above scheme, refers to the usual outsourcing of services to external experts. While it is clear that expertise, beyond that available in contracting authorities, is necessary to complete these steps, the danger is that project coherence is lost. The huge number of people involved in conception, design, implementation and evaluation implies a loss of knowledge at each step. In other words, experts contracted to develop terms of reference may not be aware of the priorities and strategies selected by other external experts at the stage of programme conception. This often results in omission, overlap or mismatch between the stages of the programme.

The bigger the programme and the more people involved, the more critical is the need to ensure overview by the project owner. The amendments introduced in the Phare implementation mechanism (with the Decentralised Implementation System [DIS], introduced in 1997) were triggered by the recognition that the Commission, i.e. its single project manager, had often not been able to follow up programme progress in the required detail. This

was due to the high number of projects managed by an individual programme officer, the large budgets attached to each of these projects, the high number of subcontracts (and related production of terms of reference, tender procedures, selection procedures and implementation monitoring.

However, maintaining a project overview has also proven difficult for institutions in the recipient countries. Similar problems arise in terms of high staff turn-over, limited human resources and management capacities and expertise to assess the value of activities.

Figure 3: **Traditional EU policy cycle**



4. Research agenda and recommendations

4.1. Research agenda

In Annex 1 of this study the authors summarise the possible criteria for evaluation. This list is not complete and not every criterion necessarily applies. However the criteria underlines that the aspects to be considered before, during and after a programme for evaluation are wide and diverse. The authors agree with the ETF (Viertel et al., 2004) that there is no single, ready-made toolkit for evaluation and consider that methods and standards need to be developed.

The following issues are considered as priority for further research:

- (a) methods for commissioning and negotiating evaluation assignments: which procedure and format is most favourable to comprehensive evaluation?
- (b) impact assessment of evaluation exercises in general: how can the adaptation of policies be facilitated by the evaluation?
- (c) the difference in impact of different evaluation methods: which methods are more effective in terms of evaluation accuracy and impact?
- (d) evaluation dissemination tools and their outreach and impact on stakeholders: which dissemination tools are appropriate for dissemination of evaluation?
- (e) evaluation dissemination methods and their outreach and impact on stakeholders and the public: which evaluation methods increase the impact and effectiveness of evaluations?
- (f) intercultural communication and their implications for international evaluation exercises: what is the 'evaluation culture'? Is evaluation an element of some cultures or can cultures of evaluation vary between them? How is the word evaluation understood in different cultures?

4.2. Recommendations

Evaluations are single tasks (or assignments/ contracts) but also have to be seen as segments

of a bigger process of reform. General acceptance of, or hostility against, evaluations are key factors in increasing or decreasing their use and level of impact. In some cases there were no *ex-ante* evaluations, in others the evaluation was purely summative and neglected the programme process and the major stakeholders involved.

In order to apply the findings from this analysis, the author team recommends the following approaches for optimal effectiveness of evaluation processes. In addition to the criteria for a complete evaluation cycle, these recommendations might also be used as a checklist before commissioning evaluation assignments.

- (a) A process oriented approach
Evaluation as a principle is based on two complementary approaches: justification of expenditure and determination of outcomes and outputs. While both dimensions are important, evaluations should attach primary importance to the latter: evaluation is understood as a development process in which the extent and quality of individual and organisational learning is analysed.
- (b) A stakeholder-oriented approach
All implementers, managers, beneficiaries and stakeholders should be involved in the evaluation exercise in order for learning to take place at all levels and among all concerned. Only by motivating people and institutions and engaging them in the evaluation, will the improvements generated by an evaluation take place.
- (c) A transnational approach
Different research traditions and diverse disciplinary backgrounds are implicit in international evaluations. Sensitivity to cultural differences, suspicion of evaluators or fear of the evaluation consequences need to be appreciated and addressed accordingly. European evaluators must be aware of the specific requirements of truly 'European' research and have considerable experience in such activity.
- (d) A policy-oriented approach
Evaluation should not be understood as a purely academic exercise, but should also

aim at partial findings for EU and national/regional (implementation and evaluation) practices in education and training. Rather than looking exclusively at the content and methodology of delivery of education, the relevant experience (successes and failures) and the effectiveness of programme/project activities should also be examined in selected case studies. Evaluation also needs to target policy-makers in terms of format, style and message.

(e) A focused approach

Although possibly contradictory to the above statements about the need to analyse the environment and broader group of stakeholders, any evaluation must define its focus. The evaluation of all and everything is impossible and would result in superficial findings. There should be clearly defined objectives and a clear agreement with the 'paymaster' on issues to be investigated. Finally, the general scope of the evaluation 'justification' or 'improvement' also should be decided. Some of these options exclude others and cannot be used contemporaneously.

(f) A sustainable approach

Evaluation findings and recommendations should be requested and produced in several formats. As with the dissemination of programme results, the distribution of evaluation results to all involved should be obligatory in order to open opportunities for learning at all levels. Evaluation products should cover the whole range of communication and technology available (reports, briefs, electronic mailings, workshops, www-sites, meetings, etc.). The evaluation should aim at the learning educating project implementers, beneficiaries and sponsors and at having a long-lasting impact.

(g) A shared-responsibility approach

Given the frequent interruption of the chain of responsibilities in major policies and programmes, the potential for achieving more coherent project steering with permanent monitoring, overview and responsibility are explored.

It is recommended that the terms of reference are renegotiated with the beneficiaries. Evaluators may yield better results in this exercise. Quality improvement can be achieved through detailed

and comprehensive agreement on programme objectives with major stakeholders.

There is a need for detailed and comprehensive agreement on evaluation objectives with all major stakeholders. In all our cases there was no systematic approach to how the participation and involvement of stakeholders was measured. The authors recommend the assignment of one project director to oversee the project and the evaluation to overcome the problem of responsibility in the programme cycle. One person would acquire a thorough knowledge of the project and as the person responsible would be the main point of contact for external evaluators. This person would not, of course, be the only point of contact for the evaluator.

Ex-ante, accompanying and final evaluations need to be exercised by external agents. During the experts meeting at Cedefop in Thessaloniki the independence issue of evaluators was extensively discussed. It was agreed that only the independent outside view of external evaluators can provide critical and objective results for summative evaluations. Exclusive internal evaluation is often not accepted as the results of the evaluation can be seen as too biased.

The evaluator should be present during all steps of project implementation. Quality improvement can only be guaranteed by sufficiently high evaluation budgets. Too low budgets result in superficial evaluation procedures (up to 5 % of project value may be necessary for comprehensive evaluation processes).

4.3. Evaluation as part of implementation

In the context of the Phare cases it became clear that most of the programmes were performed without clear methodology and guidelines for monitoring and evaluations. Monitoring procedures were basically innovated by those in charge of programme management for practical management needs. The procedures were often imperfect and did not keep clear records in accordance with specific monitoring indicators, which normally serve as an important evaluation tools.

Evaluations as such were never an integral part of programme implementation. This has also financial implications, with no budget reserved for

evaluation activities. Over several years, similar programmes were implemented in succession without the appraisal of evaluation results of the previous programme (or its cycle) but this does not mean there were no modifications or improvements from one programme to another. Such processes were intuitive and straightforward, a form of 'learning by doing'. Only after a number of years and performing an increasing variety of programmes and with emergence of the issue of absorption capacity, did it become clear how important clear indicators and good records were. This was largely influenced by pressure from the European Commission and by linking Phare assistance to structural funds procedures. In line with EU Member States, CEECs slowly started to develop evaluations in accordance with

the methodology provided for the structural funds evaluations used also for other programmes, including evaluation of implementation of national policies.

Evaluation culture depends on awareness of the positive aims and the added value of evaluations. This suggests the need for publicity in EU countries and CEECs to stimulate interest among the professional classes. As with many issues, evaluation culture needs broad social backing, not only from the State, public and international authorities but, even more, from the civil society. The establishment of professional associations in the field of evaluation may significantly advance the evaluation culture. Inclusion of various national experts into the European evaluation networks and societies may give added impetus.

List of abbreviations

BG	Bulgaria
CEEC	Central and eastern European country
CZ	Czech Republic
DG EAC	Directorate General for Education and Culture
DIS	Decentralised implementation system
ECTS	European credit transfer system
ENIC	European network of information centres
ENQA	European network for quality assurance in higher education
ETF	European Training Foundation
LdV	Leonardo da Vinci
LLL	Lifelong learning
LTA	Long-term technical assistance
MoE	Ministry of Education
NARIC	National academic recognition information centre
NCVQ	National Council for vocational qualification (UK)
NTF	National training fund (CZ)
PMU	Project management unit
ŠIOV	<i>Štátny inštitút odborného vzdelávania</i> [State Vocational Education Institute] (SK)
ŠPU	<i>Štátny pedagogický ústav</i> [National Institute for education] (SK)
SR	Slovak Republic
UVET	Programme component of Bulgarian Phare: upgrading VET
VETERST	Vocational education and training; education; research, science and technology (programme)
VÚPSVR	<i>Výskumný ústav práce, sociálnych vecí a rodiny</i> [Research Institute of labour, social affairs and family] (SK)

Annex 1 – Criteria for VET evaluation

This list of common criteria for VET evaluation was agreed on the basis of a first analysis of the case studies and after discussion with colleagues at the February 2002 Workshop in Thessaloniki

Clarification of purpose of these criteria:

We looked for criteria that could be applied to structure relevant information gleaned from the evaluation of programmes under our remit.

The timescale used to group these criteria may be appropriate for larger evaluation exercises. After considering comments from others involved in the third research report these criteria were applied to our sample of evaluation processes and reports. While we are concerned about impact evaluation we need to take into account the process of evaluation (inc. formative evaluation) if we are to make useful conclusions about evaluating impact and the design of reform programmes.

Before the implementation of the programme

1. Reform setting: what was the setting (political, social, economic, cultural) for the reform? How much attention was given to it when designing the evaluation? How important are these settings likely to be?
2. Reform aims: what was the aim ⁽²⁰⁾ of the reform? What was to change (e.g. learning, institutions, frameworks and certification)? What was to be maintained? Was all this considered for evaluation?
3. Reform actors: who was pressing for change? Who was opposed to it? How did they act? Were they motivated to participate? Why were they successful? Was a stakeholder analysis part of the evaluation?
4. Policy relevance: what was the extent of, and attention to, any explicit link between national

economic goals and the reform (e.g. skills needs, workforce development, and partnership)?

5. Policy description: was a description of national policy used to underpin the reform and its evaluation and was it intended to measure impact in terms of this? If there was no policy description what was the point of the evaluation?
6. Policy involvement: was there evidence of policy level involvement? Similarly was there evidence of bottom-up demand for reform? Did evaluation investigate this?
7. User impact: impact needs to be measured by changes in opportunity and practice amongst end users (learners, communities, organisations and countries). Was there evidence that there was sustained focus on increasing opportunity and developing practice amongst end users and how did the evaluation measure it?
8. Donor commitment: is there a declaration of intent and continued commitment from any donor authority? Did the evaluation investigate it?
9. Risk management: what did people see as the risks associated with each reform? What was done to reduce risk? How was risk management organised? Did the evaluation include programme risks?
10. Reform management: who managed the reform (people, agencies)? How well did they do? How did they see the reform? Were all actors considered in the evaluation? Was there evidence of policy level involvement? Similarly was there evidence of bottom-up demand for reform?
11. Legal embedding: is there evidence in the evaluation that, where appropriate, legal arrangements were considered as a preparatory step towards implementing the reform?

⁽²⁰⁾ Each feature of planned change will need to be broken down to reach the heart of the reform, for example, if improving learning was the aim we will need to look at relevant views of learning, predominant types of learning (formal, informal, non formal), exemplars of desired learning styles.

12. Programme handbook: was there a handbook or manual (or other material) describing the reform? How did any donor institution contribute to this publication? Were the definitions of the fields involved clear? Was the manual evaluated in terms of content?
13. Preparatory studies: were feasibility studies carried out or ex-ante evaluations planned? Did the evaluation look at the quality of these studies?
14. Social partners: was there a plan for full engagement of social partners? Were they involved in the evaluation?

During the implementation phase

15. Dissemination practice: how was the proposed reform publicised? Who discussed it? Did evaluation thematise dissemination throughout programme duration?
16. Management structure: what was the basic structure of the reform (e.g. legal, structural, management, funding)? Did evaluation consider its appropriateness?
17. Stakeholder motivation: how were people and organisations motivated to participate? How was this motivation differentiated and assessed?
18. Stakeholders' activities: how was the involvement of individuals and key groups monitored? How were they used as indicators of success or failure in programme and the evaluation?
19. Transitional arrangements: were there transitional arrangements in place during the period of reform? How were these introduced, managed and dismantled? Did evaluation consider this? How was the involvement of individuals and key groups monitored? How were they used as indicators of success or failure?
20. Mainstreaming provisions: is there evidence that the pilot phase was useful because it linked strongly to systemic change and was effectively disseminated? Were issues of mainstreaming and sustainability addressed during the expansion phase and did evaluation consider these?
21. Complementary activities: is there evidence of organisations at national, regional and local levels working jointly towards common

targets in a complementary top-down and bottom-up approach and were they involved in the evaluation?

After the programme had been completed

22. Programme context: how important was the political, social, economic and cultural influence setting for the success of the reform and how did evaluation cater for it?
23. Programme timing: on what timescale was the reform planned? How did this compare with the outcome according to the evaluation?
24. Programme budget: what did the reform cost? What were the intended costs?
25. EU priorities: how are the EU-wide common themes for VET (such as improving access to VET and raising participation) addressed in the programme and the evaluation?
26. Programme spill-over: is there evidence that the programme and evaluation practices (introduced from the EU) resulted in activity in a Member State and a programme of national action?
27. Monitoring tools: were monitoring and evaluation tools in place to identify, take account of and measure the impact of policy changes during the course of the reform and did evaluation analyse them?
28. Programme flexibility: if an *ex-ante* evaluation process was in place, did this evaluation take the possible future change of the setting into account?
29. Systemic impact: is there evidence of a systemic link between the outcomes of the reform and policy development and how was it measured?
30. Dissemination phasing: is there evidence and evaluation findings' that the dissemination process is culminating at a certain point after the beginning at the outset of the activity?

The evaluation process

31. Development process: was there evidence that the evaluation was understood as a

- development process aimed at capacity and institution building and creating self-evaluation skills amongst beneficiaries?
32. Evaluation culture: how was the evaluation process perceived by the major stakeholders (EU funders, national government, major actors, social partners and the evaluators themselves)?
 33. Evaluation method: was there evidence or suspicion of methodological weaknesses in the evaluation programme, e.g. sampling, instrument design, indicators, time and constraints?
 34. Evaluation completeness: is there evidence of the whole process of reform being fully evaluated? Were there shortfalls in evaluation effort in specific phases?
 35. Evaluation indicators: Is there evidence of use of quantitative and qualitative indicators being developed for the evaluation process? If there were no indicators (developed through benchmarking and target setting for example), what effect has this had?
 36. Evaluation skills: was there evidence of a lack of experience of evaluation practices, methods and mechanisms in any stakeholder organisation?
 37. Evaluation censorship: was there any evidence of evaluation findings and reports being shelved or not disseminated?
 38. Evaluation budget: what is the size of the evaluation budget in relation to the activity as a whole (as a rule, 5 % of budget for small projects and 3 % for large programmes should be made available)?
 39. Evaluation dissemination: to what extent are evaluation results being made available to the project partners (partner countries) and to all member countries?
 40. Reporting formats: was the dissemination material targeted to different audiences as opposed to a one-size-fits-all policy?
- Mike Coles (12.3.2002)
Bernd Baumgartl (19.2.2002)

Annex 2 – Criteria-based analyses

The following criteria-based analyses were carried out and annexed to give the reader additional information for the in-depth study of four cases. They are not intended to be read as complete text, but to provide direct access to certain aspects of a case (according to the criteria in Annex 1) and to enable selective examination of an evaluation.

Case Study 1: evaluation of Phare VET programme CZ 93.05, VET in the Czech Republic

Before the implementation of the programme

- 1., 2. and 3. – The programme setting, aims and the actors have been described extensively in the textual analysis of Chapter 2 (see Sections 2.2.1. and 2.2.2.).
4. Policy relevance. The objective of the programme was to support the Czech Government in the modernisation of its VET as a key element of economic and social reform. However, the policy framework was missing. The Strategic Review provided valuable analysis and guidance on how the programme can significantly contribute to the reform progress. Sadly, many of the points made in this document are still valid.
5. Policy description. *The strategic review of VET* provided the strategic framework for the implementation of VET.
6. Policy involvement. Throughout the programme the involvement of the Czech MoE remained minimal. The evaluation points to this as to a major drawback not only of the programme implementation but also of the programme design. The objective of the contribution to the reform of VET did not correspond to the strategy for implementation with its focus on pilot schools and lack of systemic activities. Although the managing unit paid much attention to dissemination activities and public discussion, the lack of

receptiveness, interest and acknowledgement of results from the MoE prevented the mainstreaming of outputs to national level. Further, the programme arrangement did not involve third parties from policy level, e.g. the Ministry of Labour, which resulted in the lack of linkage to labour-market restructuring and adult training.

The bottom-up initiative proved more encouraging. There was much interest from pilot schools and cooperation with them was smooth and gratefully received. Unfortunately, the programme provides a mechanism for dissemination of pilot school's work to other schools.

7. User impact. There is no doubt that the programme improved the quality of the schools and had a major impact on their development. The aforementioned foreign study visits facilitated knowledge transfer and new pedagogical methods, teaching materials and management practices improved efficiency. The evaluation also identified weaknesses in the programme such as the lack of transfer of good practices and programme results to other, non-pilot schools.
8. Donor commitment (see Section 2.2.1.).
9. Risk management. According to the evaluation report, the two concerns anticipated in the Financial memorandum (limited availability of funds, reluctance of pilot schools) were not supported by the evidence. With the risk of lack of government commitment to VET reform, such assumption in the log frame matrix of the Financial memorandum were not apparent and it put the whole programme at risk.
10. Programme management. Responsibility for overall implementation was transferred to the ETF in 1995 under the Convention signed between the Commission and ETF. ETF was charged with supporting the monitoring and implementation of the Financial memorandum.
11. Legal embedding. The new bill of the School Act took into account some of the programme results. It was proposed to Parliament for approval several times but due to

the lack of political consensus each time it was decided to hold the document for further elaboration.

12. Programme handbook. The manual for implementation was a standard Phare manual DIS. For target groups and the wider public a regular newsletter was published and extensively distributed, there was a media campaign and many publications were disseminated and discussed throughout the programme, including the Green Paper and a Policy Paper.
13. Preparatory studies. The aforementioned Strategic Review was the starting analytical document that identified major challenges and priorities. The Review was implemented in the framework of a different programme. The Phare VET programme also conducted a feasibility study and drafted the Strategic Plan.
14. Social partners. The social partners were involved through the Advisory Task Force with an important role in the programme development. Some pilot schools managed to attract interest from social partners at the regional or local level but involvement in the process of curriculum development has generally met with disinterest. This was due to the lack of systemic provision and a lack of interest in cooperation from social partners. The evaluation hardly involved social partners in finding reasons for their lack of interest.

During the implementation phase

15. Dissemination practice. The evaluation sums up the milestones in dissemination activities, including eight thematic reports analysing the experience of pilot schools activities, three synthesis reports, a draft policy paper (Green Paper); all constituting a major publication *From pilot schools to reform strategy; outcomes of the Phare programme reform of vocational education and training*; a number of international, national and regional conferences and seminars where the findings were widely discussed; the media campaign; a discussion paper *Education for prosperity – towards the learning society* and others.
16. Management structure. The PMU installed by MoE (1994) did not start the programme effectively resulting in an overall delay and their transfer to the National Training Fund in 1996 (a non-governmental organisation set up by Ministry of Labour and Social Affairs for

managing support programmes in the field of HRD).

17. Stakeholder motivation. No clear evidence in the evaluation. Stakeholders were involved through the Advisory Task Force which also benefited from international cooperation and expertise. This arrangement was effective but a larger and more systematic involvement with the MoE and other ministries and social partners did not occur, due to the initial programme setting, responsibility allocation and their lack of commitment.
18. Stakeholders' activities. No systematic monitoring of key groups took place as this was not a part of the programme design and planning. Therefore, the relevant statistics (e.g. employability of students of pilot schools) was not foreseen. Monitoring occurred at the pilot schools that mainly benefited from the programme. The evaluation did not include measurement of the impact on individuals and target groups.
19. Transitional arrangements. Not considered in the evaluation report apart from the issue of low mainstreaming.
20. Mainstreaming provisions. Some aspects of the programme results can be developed and implemented with support from the European Social Fund. However sustainability of the programme is far from being assured.
21. Complementary activities. The Advisory Task Force had great potential but was disbanded after the programme completion due to lack of funding and support from MoE, despite the participants desire to continue cooperation. The members were interviewed by the evaluators.

After the programme had been completed

22. Programme context. Apart from the issues already mentioned, it is also important to note that MoE suffered from frequent change of ministers and key staff. The MoE largely did not accept the major outputs of the programme. The period was also characterised by the lack of constructive inter-ministerial cooperation and systemic provision for social partner's involvement.
23. Programme timing. 1994-98: The initial duration was until the end of 1997 but this was extended until 31 December 1998, because of the initial delay in implementation before

the National Training Fund took charge. Also this was to allow the pilot schools to have the full education cycle of four years in order to have comprehensive results for further analysis and recommendation.

24. Programme budget. ECU 4 million. Poor disbursement in the initial phase eventually improved after the transfer of the PMU to the National Training Fund, which resulted in the extension of the disbursement period.
25. EU priorities. The programme was designed to meet the immediate priorities of the Czech Republic during the transition to market economy and labour-market restructuring. However it also contributed to the accession preparation, especially from the perspective of later developments and a more strategy-oriented approach.
26. Programme spill-over. No evidence.
27. Monitoring tools. Phare monitoring mechanisms are focused on implementation and management rather than contextual policy development. The instruments for adjustment were not in place. The programme in any case was implemented in such a difficult policy environment that the mere completion of the programme and its valuable outputs demonstrate the success of adjustment in itself. The evaluation report did not scrutinise change management practices.
28. Programme flexibility. No *ex-ante* evaluation. No specific needs evaluation was undertaken. This is a major drawback of the design of standard Phare programmes but *ex-ante* evaluations have now become standard in Phare.

However, in order to assess the reasons for the delay at the initial programme period, the Centre for foreign assistance initiated the evaluation of the programme, which was performed by BBJ Consult Berlin in 1995. The evaluation (1995) suggested the drafting of a Strategic Plan for the remaining period, appointing a new PMU Director, moving the PMU back to Prague, revising the composition and functioning of the Steering Committee by involving policy-makers, establishing an advisory body and monitoring and evaluating the experience of pilot schools in terms of processes and outputs. These

recommendations were implemented at the beginning of 1996.

29. Systemic impact. The programme achievements are largely due to the bottom-up initiatives and receptiveness/readiness to reform at the micro level and among non-governmental organisations and research institutions.
30. Dissemination phasing. The National Training Fund was the main initiator of activities, performed mainly in collaboration with the Research Institute of Technical and Vocational Education and involved the work of both local and foreign experts. Although the activities contributed significantly to public awareness, they did not result in the improvement of policy-level receptiveness and results mainstreaming.

The evaluation process

31. Development process. No evidence is available. It is however important to mention that the evaluation itself was of a fairly formal nature, corresponding to the standard Phare evaluation guidelines. It did not measure impact, nor did it perform any follow up study at the target group level. Although extensive interviewing took place, the evaluation team did not research the pitfalls of accepting the content-related outcomes of the programme. There is little evidence of understanding of the value of product elaborated within the programme no verification of meeting needs in the context of development and strategy. The feedback that such evaluation could bring was therefore limited, in spite of general understanding of major problems of the programme setting and implementation.
32. Evaluation culture. No information is available on the point. In general Phare evaluations were considered as a threat more than as a consulting tool. Phare arrangements for evaluations rely on contracting out to international consortia. Although these involve some local experts in the evaluation, Phare generally did not contribute significantly to the development of the evaluation culture in the country, nor did it enhance expert capacities in evaluations. Evaluation arrangements are decided within the Phare programme bureaucracy, instead of being made the responsibility of managing structures and policy-makers. Evaluations are therefore perceived as imposed externally and

naturally not seen as a constructive tool for policy planning and adjustment.

33. Evaluation method (see Section 2.2.1).
34. Evaluation completeness. All components of the programme were evaluated.
35. Evaluation indicators. The evaluation report often refers to the lack of measurement of the programme impact and lack of relevant statistics. While it is an important issue, this was not foreseen in the programme. It is however unclear why a serious attempt at impact measurement was not foreseen in the evaluation action itself.
36. Evaluation skills. The contract was assigned to the OMAS Consortium of the Middle Region. Further information is not available. The failure to measure impact with indicators suggests a limited use of evaluation tools.
37. Evaluation censorship. The recommendations produced by the evaluation have not been taken into account by key personnel to any significant extent.
38. Evaluation budget. No information is provided in the evaluation report.
39. Evaluation dissemination. It was made available to appropriate stakeholders in the programme.
40. Reporting formats. No evidence.

Case study 2: evaluation of Phare VET programme SR 94.03, VET reform in the Slovak Republic

Before the implementation of the programme

1. Programme setting. Programme SR 94.03 was a direct continuation of the VET component of the previous Phare programme Labour force market restructuring, T 91.08, designed when Slovakia was still a part of Czechoslovakia. This earlier programme had produced a general strategy for VET reform (Executive study by Birks Sinclair in 1993) and had culminated in the development of new curricula for first grade classes in 20 pilot schools, covering eight professional branches.
2. Programme aims (see Section 2.3.2.).
3. Programme stakeholders. The start-up documentation very clearly identifies those to be involved in the project: the MoE, the Pilot

Schools, educational research institutions such as the ŠPU, later to be replaced by the ŠIOV, regional inspectorates, social partners and the Ministry of Labour, Social Affairs and Family. The role of the PMU and the LTA, as fixed in the terms of reference, was to provide the necessary management following the DIS and to prepare and assist all involved in reaching the objectives of the different subprojects. Key components of the implementation were contracted by the PMU to the consultants BBJ (Germany) and Birks Sinclair (England).

The Financial memorandum for Phare programme SR 94.03 had been signed by a government that went out of office a few weeks later. The new government was not opposed to the programme but was not interested or supportive either. Its success was in the main achieved through the individual effort and enthusiasm of committed teachers and school managers.

4. Policy relevance. The objective of the programme was to support the Slovak Government in the modernisation of VET as a key element of economic and social reform.
5. Policy description. The basic documents of the programme, the Financial memorandum and terms of reference for the LTA, had been produced within a feasibility study carried out by Dutch consultants, Hobbeon and special bridging technical assistance was provided by the European Commission. The 1993 recommendations of the executive study had been adopted by the Slovak government as policy guidelines.

The following key components of the programme were agreed upon:

- (a) curriculum development in selected job families (mechanical engineering, clothing/textiles, food processing, gardening, electrical engineering, construction, economy/commercial service and transport);
 - (b) teacher training in order to facilitate the implementation of these new curricula;
 - (c) provision of learning equipment for pilot schools;
 - (d) establishment of partnerships with vocational schools in the EU;
 - (e) development of a national strategy for VET and dissemination of programme results.
6. Policy involvement. Throughout the programme the involvement of the Slovak MoE

remained minimal; no initiatives were being taken by the Steering Committee; frameworks for curriculum implementation and dissemination of the products of the Phare programme SR 94.03 were implemented only reluctantly and were fraught with restrictions. The completion of the programme and the successful developments for Slovak VET within the programme were secured by an enormous input of work and time by dedicated individuals within the management structure rather than by effective management and implementation procedures.

7. User impact. The investment made in the 20 schools heralded a new era in their life, exposing them to new pedagogical and technical practices and initiating a healthy desire for improvement, effectiveness and efficiency. Although not always supported from the programme management, the schools mobilised the competence of staff to take initiatives of their own.

The status of the schools also received a boost in their regions. They now have privileged contacts with social partners in their sector, have experience in international networking and are more aware of their potential role in the key issue of regional development. Some pilot schools expressed their intention to become fully-fledged regional knowledge centres for participation in future EU programmes.

It is at the pilot school, teacher and student levels that the programme has achieved its impact. In their view, teaching and learning has changed its concept opening people's eyes to alternative methods and is a process which is irreversible.

8. Donor commitment. The programme was part of the EU Phare programme which indicates a principal and substantial political priority from the donor. It is also conditional that an agreement of the government to the projects exists. However, at the implementation level, the monitoring and support by the EU delegation was suboptimal. This was partly due to under-staffing but also partly due to mismanagement, which had serious consequences some years later.
9. Risk management. A formal risk assessment was part of the planning documents but it did

not cover all possible risks (especially political risks), neither did the programming address the risks identified.

10. Programme management. The role of the PMU and the LTA, as fixed in the terms of reference, was to provide the necessary management following the DIS and to prepare and assist all involved in reaching the objectives of the different subprojects. Key components of the implementation were contracted by the PMU to the consultants BBJ (Germany) and Birks Sinclair (England). Both consultant and local coordinator underperformed.

Throughout the programme the involvement of the Slovak MoE remained minimal; no initiatives were being taken by the Steering Committee; frameworks for curriculum implementation and dissemination of the products of the Phare programme SR 94.03 were implemented only reluctantly and were fraught with restrictions. The completion of the programme and the successful developments for Slovak VET within the programme were secured by an enormous input of work and time by dedicated individuals within the management structure rather than by effective management and implementation procedures. In spite of some remarkable successes, the different subprogrammes did not manage to produce coherent and equally acceptable results. Results remained largely limited to pilot school level.

11. Legal embedding. No evidence of legal arrangements except for a centre which in future could act as a central clearing agency for VET reform (ŠIOV). In respect of the development of a VET reform strategy for Slovakia the programme has, at best, laid the groundwork but the concrete work still remains to be done. There were only clear statements of the new Minister that recommendations were to be followed and legal changes would be made, as indicated by the evaluators, at programme end.
12. Programme handbook. No evidence. The EU/Phare has its own Manuals for Implementation of programmes in general.
13. Preparatory studies. A feasibility study was carried out by the Dutch consultants, Hobbeon. The 1993 recommendations of the executive study had been adopted by the

Slovak government as policy guidelines. Some time had elapsed until the programme inception and the objectives were partly outdated by that time.

14. Social partners. The efforts to interest social partners and major policy-makers such as the Ministry of Labour in VET reform have largely remained unsuccessful. Some pilot schools managed to raise some, if short-lived interest from social partners at the regional or local level but involvement in the process of curriculum development has in general met with disinterest. Most of these organisational bodies are at present in a *stato nascendi* and finding competent consultants or staff for VET ranks low in their priorities. Although the pilot schools have, by comparison with other EU countries, discovered the importance of social partners, they were only able to establish sustained working relations with local firms and craftsmen's guilds at best.

During the implementation phase

15. Dissemination practice. The failure to provide for the necessary legal framework for the horizontal mode of dissemination through pilot schools has resulted in fully equipped pilot schools not being allowed to fulfil their role, thus rendering this potentially powerful tool of dissemination useless to the aims of VET reform. ŠIOV stepped in with a traditional, but inadequate, top-down model of dissemination through publications and conferences. Since the application of educational products of the other programme components was restricted to pilot schools only, this exercise remained largely ineffective. The development of a national VET reform strategy was touched upon during the final conference by two leading experts but its realisation remains a task for the future
16. Management structure. In contrast to the Financial memorandum, the PMU was permanently under-staffed, too under-paid to attract highly qualified managers and under-valued in terms of the status of the project authorising officer, who in other programmes, was at Head of Section or State Secretary level. The curricula for the first grade had been drawn up by members of the ŠPU, who had also designed Phare programme SR 94.03 under the auspices of the Ministry of Labour,

where the PMU for Phare programmes T 91.08 was situated. Development of curricula started at school level, under the guidance of ŠPU but was troubled by financial problems of ŠPU and disagreements with pilot schools.

At this stage a major change occurred which was to influence the further development of programme SR 94.03 and later threaten to destroy it altogether. In January 1996 the competences for vocational education were moved from ŠPU, where they had been for over 40 years, to the ŠIOV, a new institution created several years earlier and which, at that time, had lacked the capacity for this task. ŠPU was no longer involved. The formal competence for Phare programme SR 94.03 rested solely with the MoE, where the PMU for the programme had also been set up.

In 1996 the programme had virtually come to a standstill. There was no working PMU at the MoE and ŠIOV, as the institution responsible for vocational training, had not been contracted by the MoE for Phare programme SR 94.03. By 30 September 1996, a year before expiry, only 27 % of the ECU 4 million allocated to the programme had been contracted and only 10 % had actually been disbursed. Severe disagreements between the Slovak MoE and the EU authorities over another Phare programme aggravated the situation, which was not helped by the rapid succession of different programme managers at the ETF. Teachers and school administrators at school level were left with the task of developing the new curricula.

At this point a new procedure was negotiated by the LTA with the parties concerned and new heads of PMU and the Steering Committee installed in the MoE in October 1996. In December a new Work programme (WP3) was negotiated with the ETF. In March 1997 ŠIOV was contracted to the programme and started to coordinate national working groups to secure the implementation of all subprogrammes. This was the turning point for the project. Two core sections of the programme, curriculum development and technical upgrading of pilot schools, were immediately taken over by ŠIOV and the national working groups respectively and the final success and impact of these subprogrammes

is largely due to the dedication of the responsible experts.

The improved cooperation between the new steering committee, the PMU, ŠIOV, the LTA, the programme manager of the European Commission delegation and the ETF programme manager, resulted in an unparalleled acceleration of contracting and disbursement. By 30 September 1997, 99.9 % of the total amount of ECU 4 million had been contracted. To compensate for the earlier delays, all missing components of the different subprogrammes were now brought into operation together, often with adverse effects.

17. Stakeholder motivation. Pilot Schools were the main beneficiaries, being the main participants in this programme. In the earlier phases of the programme they were not encouraged to work, were impeded or unsupported or were unpaid for it.
18. Stakeholders' activities. No systematic monitoring took place except of evaluation sheets after training sessions.
19. Transitional arrangements. The lack of monitoring and evaluation instruments has impeded self-assessment and reengineering of the content of the programme. Initiatives almost only ever took place if in the interest of individual parties.
20. Mainstreaming provisions. Pilot schools gained competences, expertise and responsibility as a result of the programme, as did the VET Institute ŠIOV. The extension of results to the national system has not yet been initiated by the responsible authorities.
The investment made in ŠIOV is too great for the MoE not to continue to use the capacity that should exist there to carry out additional VET reform.
21. Complementary activities. Empowerment and expertise provides pilot schools with the potential to become regional knowledge centres, key-players in regional development and recipients for EU-financed programmes. However, there was little involvement for external parties.

After the programme had been completed

22. Programme context. See textual analysis.
23. Programme timing. 1994-98: the deadline for commitments was 30 September 1997, the implementation period finished on

30 September 1998. Despite delays and difficulties the programme was completed in time.

24. Programme budget. By 30 September 1996, a year before expiry, only 27 % of the ECU 4 million allocated to the programme had been contracted and only 10 % had actually been disbursed. By 30 September 1997, 99.9 % of the total amount of ECU 4 million had been contracted. The correctness of financial procedures was certified by an independent final audit.
25. EU priorities. The programme served as a means of allowing continuing reform in a difficult and volatile environment. Its completion at the start of the present government is a guarantee that none of its various products will be lost. The results of the programme have prepared the ground for a more labour-market oriented restructuring of Slovak VET, addressing the key issues of integration of schools and rationalisation of the school system. Innovation in VET is also mentioned as one of the priorities of the Accession Partnership and the Slovak intention to join the EU. A follow-up programme to valorise the investment in selected pilot schools at the national VET level is crucial.
26. Programme spill-over. No evidence.
27. Monitoring tools. No systematic monitoring took place, little transparent documentation is available and no control or supervision, from those responsible procedurally or politically, took place. Initiatives almost only ever took place if in the interest of individual parties.
The lack of monitoring and evaluation instruments has impeded self-assessment and reengineering of the content of the programme and eventually the evaluation task was hampered by the non-existence of planning instruments, a clear and logical framework (matrix), performance indicators or achievement targets.
According to the Financial memorandum, Article 2.5: Dissemination of Results, the MoE, together with the Steering Committee, was to monitor and evaluate the projects and be responsible for the dissemination of the experience from the reform projects and discussions about future developments. Further, the results of the curriculum development component were to be published

together with the strategic outline for the future development of vocational education. However, the programme was carried out and finalised without the above occurring. Throughout the implementation period 1994-98 the main parties, as defined in the Financial Memorandum, changed both in institutional terms and function. The Ministry designated a person to chair the steering committee and provided the necessary logistical support to the programme. It was not so involved in monitoring and evaluating the components of the programme, as defined in the Financial Memorandum.

Little or no information could be obtained as to minutes or documentation of decisions taken of regular meetings of the steering committee, the advisory task force, the national experts responsible for work of the national working groups, the five national working groups nor of the meetings with all the headmasters of the pilot schools or meetings of members/coordinators of the national experts group and/or national working groups with the respective school coordinators. There are two final reports, one prepared by the LTA the other by experts contracted by PMU. Neither was discussed and approved by the steering committee of the programme. Staff changes within institutions occurred during the project which sometimes were aimed at transferring responsibilities but without regard to experience, expertise or even funding (the shift in 1996 from ŠPU to ŠIOV). The MoE and the management team had not used authority of the DIS to prolong the programme and for the last two years it became largely a technical exercise.

28. Programme flexibility. No *ex-ante* evaluation. An interim evaluation was carried out but except for the school level, it was unsuccessful in turning the project around due to political barriers.
29. Systemic impact. See textual analysis (programme assessment).
30. Dissemination phasing. See textual analysis (Results and impact).

The evaluation process

31. Development process. Great importance was attached to the evaluation process proceeding in an inter-active way, assigning a

significant role to the available central and eastern European and Slovak expertise, thereby hoping to stimulate an intensive national discussion on the Phare VET programme and its value for future VET policy-making. Also a number of recommendations for the future were made.

32. Evaluation culture. Slovak expertise came from various people involved in the project, i.e. representatives of the MoE, the PMU, VET teachers and administrators from pilot schools, specialists from ŠIOV and independent national external experts who could provide an objective view on the project notably from the Slovak Academy of Science, from the VÚPSVR, the ŠPU and the Phare coordinator from the Slovak Government office. The Slovak national observatory on VET and labour market helped with management issues and logistics in Slovakia and was responsible for the organisation of the Programme review conference that took place on 5 and 6 February 1999. Evaluation was definitely and perhaps with reason perceived as a threat by part of the project management.
33. Evaluation method (see Section 2.3.2.).
34. Evaluation completeness. All components of the programme: curriculum development, teacher training, equipment and learning materials, EU-partnership and study visits and development of national strategy and dissemination were considered, evaluated and recommendation made.
35. Evaluation indicators. The evaluation process was based on the analysis of the following source materials:
 - (a) the central monitoring documents of the Phare programme;
 - (b) the mid-term evaluation of 1997;
 - (c) final reports of PMU and LTA;
 - (d) reports produced within the programme;
 - (e) analysis of the programme documentation at the ŠIOV;
 - (f) visits to pilot schools by the evaluation team;
 - (g) workshop reports of the Programme review conference held on 5 and 6 February 1999;
 - (h) analysis of questionnaires for pilot-school members and non-pilot school teachers;

- (i) inputs by independent Slovak educational experts.
- 36. Evaluation skills. The composition of the evaluation team was international, inter-disciplinary and wide-ranged. It included experts from the Irish organisation Farrell, Wegimont and Associates, who had already carried out the mid-term evaluation of curricula of this Phare programme in 1997, an experienced evaluator from the Netherlands, an Austrian educational researcher specialised in central and eastern Europe, a VET reform analyst and a PMU deputy Director of the VET Phare programme in Bulgaria. Slovak members of the evaluation team came from the Vocational Information Resource Centre in the VÚPSVR and the Sociological Research Institute 'Focus'.
- 37. Evaluation censorship. The evaluators were not allowed to send the report to stakeholders or project participants and it looked as though the report had been shelved. Two years after the evaluation an OMAS team investigating fraud uncovered this report and used it in their work.
- 38. Evaluation budget. EUR 49 000 for a programme of ECU 3 million, hence too limited a budget.
- 39. Evaluation dissemination. See above.
- 40. Reporting formats. The format of the report envisaged targeted recommendations at different levels of participation and they were delivered orally.

Case study 3: evaluation of Phare VET programme BG 96.04 upgrading VET in Bulgaria within vocational education and training; education; research, science and technology (VETERST)

Before the implementation of the programme

1. Programme setting. See introduction of textual analysis.
2. Programme aims (see Section 2.4.2.).
3. Programme actors. The start-up documentation clearly identifies the institutions and the main parties to be involved in the project: the

Ministry of Education, Science and Technology, the national Working groups composed of Ministry officials and representatives from pilot schools, pilot school teachers and administration, regional Inspectorates, social partners and the Ministry of Labour and Social Policy. The role of the PMU and the LTA, as fixed in the terms of reference was to provide the necessary management following the DIS and to prepare and assist all those involved in reaching the objectives of the different subprojects.

4. Policy relevance. The objective of the programme was to support the Bulgarian Government in the modernisation of its education and training system as a key element of economic and social reform and to ensure the availability of national human resources for economic and social re-structuring.
 5. Policy description. Programme BG 95.06 was a direct continuation of the VET component of the previous Phare programme BG 93.14-02 (Post-secondary vocational training project). This earlier programme had produced a general strategy for upgrading VET and had developed relevant practice and expertise.
 6. Policy involvement. A government that was soon to go out of office had signed the Financial memorandum for Phare programme BG 95.06. The programme was kept on course, despite frequent government changes at the outset, by the LTA and PMU and since May 1997 the permanent programme authorising officer and dedicated officials in the Ministry's VET Department.
 7. User impact. The programme in question has been designed and implemented on the assumption that the 'pilot' and 'feasibility' approach would lead to an effective implementation and mainstreaming of results at national level. As in UVET, a mainstreaming of the whole programme throughout the Bulgarian education, training and Science and Technology systems beyond the 30-odd pilot schools could be set in motion through a system of dissemination, new curricula, modules, teacher training methods and methodological expertise.
- It is perhaps at the level of pilot schools, administrative staff, teachers and students

that the programme has achieved its strongest impact. In their view, teaching and learning has changed its concept and opened people's eyes to alternative methods and is a process which is irreversible.

8. Donor commitment. The programme was part of the EU Phare programme which indicates a principal and substantial political priority from the donor. It is also conditional that an agreement of the government to the projects exists.
9. Risk management. A formal risk assessment was part of the planning documents but it did not cover all possible risks (especially political risks), nor did the programming address the risks identified.
10. Programme management. The role of the PMU and the LTA, as fixed in the terms of reference, was to provide the necessary management following the DIS and to prepare and assist all those involved, in reaching the objectives of the different subprojects.

Most programme activities proceeded according to the Work programmes. However the UVET programme only completed work on 18 professions instead of the planned 20/25. Dedicated managers and school staff achieved the completion of the programme. All but one of the Teacher career path centres failed to achieve sustainability and the technological unit in Kostimbrod failed. The limited dissemination and mainstreaming of pilot curricula to non-pilot classes and to non-pilot schools was due to the problem of the Teacher Career Path and this is a reflection of the wider structural issues in teacher training.

11. Legal embedding. The new VET law was envisaged to be in place during the programme but it failed to be passed by parliament. It was, however, adopted soon after project end.
12. Programme handbook. No evidence. The EU/Phare has its own manuals for implementation of programmes in general.
13. Preparatory studies. An EU expert group had written a feasibility study but it was outdated at programme inception.
14. Social partners. Efforts to involve national social partners and major policy-makers such as the Ministry of Labour in VET reform have

so far been unsuccessful. It is at the local level in many municipalities and communities that this relationship works. However the result of this participation exercise has had positive results. Even if not effective, both social partners and governments have become more familiar with the concept and the social partners understand their potential role in the development and implementation of VET policy. Education, especially continuing, training, is a field where social partners in the EU are taking a leading role and greater interest.

During the implementation phase

15. Dissemination practice. In general, dissemination activities as far as they were executed during the programme have not yet led to a permanent mechanism for circulating and exchanging information. There is a role here for the institutions set up under the programme, which are well equipped to take over this task (see also Section 2.4.2.).
16. Management structure. The programme documents envisaged the establishment of a steering committee to make strategic decisions on the implementation of the programme. The composition of the steering committee, as described in the Financial Memorandum, comprises representatives of the Ministry of Education and Sports, Ministry of Labour and Social Policy, Ministry of Economy and social partner organisations. It proved difficult to set up the Steering Committee, it was a new idea in Bulgaria and took until 1998 to achieve, a considerable delay. A suggested committee structure was provided by the ETF but greater pressure from the EU delegation could have helped speed the process. Initially there were 46 proposals for membership of the Steering Committee that were reduced to 18 for practicality. This committee was the first time groups such as employers, teachers unions and ministries had come together in one forum. After the second meeting their role was more fully understood and interest was taken in the functioning of the programme, checking the results of working groups. Some cooperation of this kind also spread to pilot areas.

Notwithstanding the sharp learning curve and the occasional lack of transparency in deci-

sions on procurement, the largely successful implementation is due to the combined management ability of the PMU, predominantly competent consultants and the management style of the programme manager at ETF. The PMU has been a valuable training platform for a number of enthusiastic young Bulgarian managers and staff, who developed significant competence in EU programmes and procedures. This resulted in the creation of such new institutions as the human resources development centre and the LdV and Socrates office, thus successfully contributing to the improvement of Bulgarian education management. However, while supportive to the programme, policy-making and mainstreaming of pilot results has so far been less successful and not all components were executed as well as envisaged.

17. Stakeholder motivation. A continuing problem in Phare activities is the inability to pay State employees with project money. Consequently teachers and administrators can only rely on their salaries. While this may be justifiable in the EU, low teachers' salaries often do not provide enough incentive for the additional work generated by a programme. The award of diplomas does attempt to provide an incentive but the failure to recognise teacher training as a means of career progression was criticised by the evaluators.
18. Stakeholders' activities. As mentioned, the PMU took over substantial project monitoring and evaluation activities. The management of this project was successful in these respects.
19. Transitional arrangements. The evaluators attached much importance to the lack of transitional provisions and provided detailed recommendations to the responsible parties on how to address this need.
20. Mainstreaming provisions. The absence of a legal framework for mainstreaming the experience of pilot schools inhibited dissemination. Apart from information packages for non-pilot schools in UVET, the top-down dissemination strategy adopted by the consultants did not help this situation and the opportunity to provide support for this process through the Regional Inspectorates was not fully used. Moreover, the financial crisis after the programme resulted in less attention from the

Ministry of Education and Sports. The impact of the programme activities on key stakeholders was further weakened by changes of some of those involved, for example trained administrators, experts or teachers were substituted or moved to other better paid jobs. The new VET Law, although drafted with suboptimal input by the Ministry of Labour and Social Policy and the new NAVET agency provide concrete opportunities for the future.

21. Complementary activities. As a result of the programme, three teacher training centres, three foreign language centres, a national technology information centre and an information management centre emerged. Expertise was passed from the PMU to the emerging and sustainable human resources development centre, the Socrates office, LdV and other European integration facilities, thus successfully contributing to the capacity and institution-building in Bulgarian education management.

After the programme had reached maturity, been completed or was terminated

22. Programme context (see Section 2.4.1.).
23. Programme timing. The programme lasted from 26 June 1996 until September 1999. After difficulties during the Work programme (WP1) and repeated delays in the first stage of the delivery of equipment, by 1997 the programme proceeded well on schedule. The cooperation between the PMU, the national working groups, the LTA and the programme managers of the EU at the European Commission delegation and at ETF, although not completely free from friction, compensated for the delayed set-up of the Steering Committee (mid 1998) and resulted in a smooth procurement and expenditure.
24. Programme budget. At the end of the disbursement period, 99.99 % of the total amount of ECU 9 million had been spent. A few subprogrammes needed an extension of six to nine months to finalise disbursement (UVET, foreign language training, financial management for school education and the higher education evaluation and accreditation agency).
25. EU priorities. At the end of VETERST Bulgaria already has 60 % of the *acquis communautaire* in the education field. As a result Bulgaria was beginning accession negotiations with Brussels in the field of education.

Collaboration with foreign experts will naturally continue and, as happens in the EU, has shifted from a 'foreign expert-recipient' relation to a partnership of colleagues at peer level. All these are features of a vibrant and flexible system of continuous adaption of the education system to a changing market environment. Whilst Bulgaria has not succeeded in establishing a ready and definitive system, this is not a failure as it was an impossible goal to start with.

26. Programme spill-over. In terms of individual learning and organisations involved in the programme, many results could be observed. The systemic learning at Ministry level was limited. Some dissemination conferences were attended by experts from neighbouring countries (Macedonia, Romania, etc.).
27. Monitoring tools. As required, regular reporting was undertaken and made available to the evaluators who acquired much of their first understanding on this comprehensive information.
28. Programme flexibility. No *ex-ante* evaluation had been carried out, instead a feasibility study was carried out (see above).
29. Systemic impact. Vocational training is linked to the labour-market survey. The Deputy Minister for Education and Sports, Mr Panchev, wrote to regional inspectors asking for information on employment. There will be a national information system. All schools will be asked for information at the beginning of the next school year about which students obtained jobs. This survey was implemented together with the National Statistical Institute and local structures of the Ministry of Labour and Social Policy, the regional employment councils. Previously, the Ministry of Education and Sports would tell schools how many places and in what subject were required but now schools make proposals to the regional Inspectorates for their future provision. Moreover, each of Bulgaria's 28 districts has a development plan. Schools negotiate on this basis with local government and employers. Depending on demand for places the school approaches the Ministry of Education and Sports for funds. Experts from regional Inspectorates check with local employment offices whether estimates of

vocational schools about demand are correct. A clear demonstration of success is that 80 % of graduates have found a job or started their own business.

Vocational schools now have passports but they still need to be restructured. It is planned to reduce the existing 482 schools (plus 75 classes in general secondary schools) to 478 but closing schools has impact on local employment.

The Ministry of Education and Sports considers that there is now joint responsibility with the Ministry of Labour and Social Policy for vocational training. The Ministry of Labour and Social Policy however, paints a picture of non-consultation and the removal of social partners from vocational training bodies as a result of Ministry of Education and Sports prepared legislation. However the Ministry of Labour and Social Policy has had long term experience of working and consulting with social partners. At the programme valorisation conference project authorising officer Totomanova, for programme BG 95.06, characterised the situation as 'the relationship with the social partners has moved forward', which is a fair judgement.

30. Dissemination phasing. It is apparent from all interviews outside the capital that schools, inspectorates and higher education institutions feel that they receive insufficient information and demand more. The evaluators feel that the Ministry should include dissemination in the task list of these and the Regional Inspectorates. To begin with, however, most information must leave the Ministry itself in a transparent way.

The evaluation process

31. Development process. Great importance was attached to the evaluation process proceeding in an inter-active way, assigning a significant role to the available central and eastern European and Bulgarian expertise, thereby hoping to stimulate an intensive national debate on the Phare VET programme and its value for future VET policy-making. Also a number of recommendations for the future were made.
32. Evaluation culture. The people involved in the programme were receptive to the idea of evaluation. Schools sometimes confused the evaluators with 'inspectors'. Critical remarks were,

however, met with resistance. The Programme review conference on 14 February 2000 brought together for the last time all the main parties involved and helped to test the initial findings and theories of the evaluators. It also resulted in a public commitment by the MoE to implement the evaluation team's immediate and urgent recommendations.

33. Evaluation method. As an inherent feature of the programme design, the evaluation struggled with the difficulty of measuring seven diverse subprojects with one common approach. However, a variety of tools was used to understand project achievements and impacts.
34. Evaluation completeness. All components of the programme were evaluated: UVET, teacher career paths, foreign language training, financial management for school education, development of science and technology, school drop-out in general education, a higher education evaluation and accreditation agency, a science park feasibility study and the PMU.
35. Evaluation indicators. The evaluation process was based on the analysis of the following source materials:
 - (a) the central monitoring documents of the Phare programme;
 - (b) the intermediate assessment of OMAS (as part of a general assessment of all Phare activities);
 - (c) final reports of PMU and Consultants for all subprogrammes;
 - (d) reports produced within the programme;
 - (e) analysis of the programme documentation at the PMU and pilot schools;
 - (f) interviews at and visits to the Ministries of Education and Labour, pilot schools, regional authorities and institutions set up during the programme by the evaluation team;
 - (g) the work groups reports of the Programme valorisation meeting on 14 February 2000;
 - (h) analysis of questionnaires for programme recipients;
 - (i) inputs by independent Bulgarian educational experts.
36. Evaluation skills. The composition of the evaluation team was international, inter-disciplinary and wide-ranging. Bulgarian members of the evaluation team came from the New

Bulgarian University, from the Economic University and the National Institute of Education. The team also included the Austrian evaluator of the Slovak Phare VET programme, an experienced VET university teacher and an evaluator from the UK, a VET reform analyst from Sweden and a British NGO Director resident in Bulgaria while implementing other EU programmes in Bulgaria.

Further Bulgarian expertise came from people involved in the project, i.e. representatives of the Ministry of Education and Sports, the PMU, VET teachers and administrators from pilot schools, specialists from research institutes and independent national external experts who could provide an objective view on the project, notably from the Bulgarian Academy of Science, from the National Employment Institute and the National Institute of Education. As detailed in the contract with the ETF, the Bulgarian national observatory on VET and labour market (which had now become a part of the human resources development centre) helped with organisational issues and logistics in Bulgaria and was responsible for the organisation of the 'Programme valorisation meeting' which took place on 14 February 2000 in Sofia.

37. Evaluation censorship. No signs of censorship, only the comments on the failed project component in the technological unit were not well received by everyone. However, the evaluation results were made public.
38. Evaluation budget. EUR 49 000 out of EUR 9 million, therefore a small allocation (although one third of budget was simply for procurement).
39. Evaluation dissemination. Substantial dissemination products were produced by the PMU and meetings were held to share project experiences. While some programme components are bound to stay as institutions or national policies, sustainability of some subprogrammes is far from guaranteed. If no additional donor support steps in, it will depend exclusively on the decisions actions, of the MoE whether some products can and will be used further. Public statements by the Minister and his Deputies are encouraging but; the prospect of concrete measures and funding for implementation of mainstreaming is still bleak.

40. Reporting formats. One final workshop and one report were requested as evaluation products.

Case study 4: transnational analysis of NARIC. External evaluation of the NARIC network

Before the implementation of the programme

1. Programme setting (see Sections 2.5.1 and 2.5.2.).
2. Programme aims (see Sections 2.5.1 and 2.5.2.).
3. Programme actors (see Sections 2.5.1 and 2.5.2.).
4. Policy relevance. The national centres are designated by the participating countries. They are responsible for providing institutions and the public with information about higher education qualifications, with a view to facilitating the recognition of such qualifications in participating countries for academic and, in many cases, economic purposes. A strong transnational dimension with the focus on foreign education systems and qualifications has shown a remarkable stimulus for the economy. National institutions have clearly recognised this factor.
5. Policy description (see Chapter 2).
6. Policy involvement. Policy analysis results are only appreciable from a qualitative perspective for the 'other' areas included by some of the respondents, because of the small number of answers. The following indicators emerge from the crossed analysis of the resulting tables:
 - (a) academic recognition is the area of major interest to governments and other policy-makers;
 - (b) professional recognition is the second area of major interest to governments and other policy-makers;
 - (c) VET is an area where fewer recognition centres are active because of their mandate;
 - (d) informal qualification is an area where the number of recognition centres active is almost identical to professional recognition.There are signs of interest of national governments and other policy-makers in other

areas. In general, frequency of contacts and interest tend to coincide, which can also be regarded as an indicator of the acknowledgement of the role played by recognition centres in these areas.

7. User impact (see Chapter 2).
8. Donor commitment. Although the situation varies from country to country, recognition centres are mainly, largely or completely financed by their national administrations. Since its proposal for the establishment of the network, the Commission made clear that costs for staffing and running of the information centres were to be charged 'against national budgets, where such units do not already exist'.

The European Commission finances some project activities carried out by the recognition centres with a yearly budget of its Socrates II (previously, Socrates; previously Erasmus) programme. Funds, which are managed by DG EAC/A2, are granted to single recognition centres or, more rarely, to groups of them on a competitive basis.

According to the definitions developed in the second report on the reform of the Commission from 1999, these funds are to be considered as subsidies (or grants).

9. Risk management. The study points out in one of its recommendations that there is the risk of neglecting the European dimension of the network. More visibility and dissemination of materials and exchange of lecturers could reduce the tendency of duplicating activities, abuse or ignoring budget limits.
10. Programme management. The management level was analysed from a sustainability perspective: the establishment year of the present management of the recognition centres has been observed against the mother organisation's seniority of experience in academic recognition for each different nation. The large majority of their mother organisations practiced academic recognition well before the appointment of the present recognition centres.

For guaranteed future quality management an interesting remark was made in one answers of the textual questionnaires: 'I would like to introduce a yearly (or 2-yearly) recognition procedure test, which will lead to official

NARIC network practical recommendations concerning the process of credential evaluation aiming at abolishing the existing bureaucratic barriers for mobility in the Member States. The test comprises submitting the same 5-10 (fake and hidden) recognition requests to each NARIC. It will detect the shortfalls in management and administration by checking the occurrence of “good practice” measures.’

11. Legal embedding. The legal framework of NARIC has been described extensively in the evaluation report. The majority of recognition centres are of a public nature. Therefore, it was of little surprise that the modalities of appointment of the majority of them reflect typical public approaches, i.e. internal delegation of powers, ministerial orders, normative acts, direct agreements or direct appointments. These are modalities of appointment that, by their nature, exclude competition. It is surprising therefore that the percentage of recognition centres being appointed with modalities not based on competition (96 %) is by far larger than the 71 % of those recognition centres of a public nature. In other words, most recognition centres not of a public nature are appointed using regulatory instruments of the public administration.
12. Programme handbook. Programme implementation took mainly place by way of individual and separate projects of diverse character. Therefore, no handbook had been developed. It was however noted in the evaluation report that reporting, monitoring and documentation are insufficient and recommendations for new tools were put forward.
13. Preparatory studies. No evidence about *ex-ante* studies was given. However nine recognition centres which were subject to formal quality control mechanisms were taken into consideration. In general, they referred to an overall assessment of recognition centres’ performances, being carried out by internal audit (four cases) and/or by external review of the mother organisation (six cases). Only three were subject to the guidance of a steering committee, while peer reviews were being implemented by one NARIC on credential evaluation. None of the recognition

centres was subject to more formal processes, such as ISO 9001 certification.

14. Social partners. The only evidence from social partner involvement (which may be less critical in higher education than in VET) is an interview with a stakeholder. It is indicative that the one organisation interviewed (Union of Industrial Employers Confederations Europe) was not aware of the existence of the recognition centres and their network; all the others knew to some degree and agreed to be interviewed.

During the implementation phase

15. Dissemination practice. The study makes critical remarks about dissemination of information produced by the programme. Dissemination of network project results is not the main concern of most recognition centres. A minority disseminate their reports to other centres and some of them confused dissemination of results with delivery of the activity report to DG EAC. The executive summary points out that the specific network obligation needs to be addressed. In particular it needs to set how and when somebody is responsible for the authentication of information, the mechanisms for collection and dissemination and tools for secure dissemination.
16. Management structure. The network structure was analysed through qualitative questions directed to NARIC national administrators. For example one of the respondents identified a major problem of the network structure: Unlike other networks, NARIC has no specific European secretariat (like Eurydice, Tempus, etc.). Only one person from the European Commission is devoted to the coordination of the NARIC and it is not their only task. Consequently there are relatively few network activities undertaken except those by some individual centres. ENIC is, according to the evaluators, a better example of what a network should be like. Finally the evaluators recommended, ‘that DG EAC should reconsider the structure of the NARIC network in function of its (network) mandate. In particular, DG EAC’s involvement and role shall be reassessed. No major resistances emerge from recognition centres in acknowledging the right of DG EAC to play a steering role of the network, which appears to the evaluators

to be a proper function for the Commission to play. This will require additional human resources from the Commission side'.

17. Stakeholder motivation. It is the nature of their activity, the mandate of many and in the interest of all NARICs to participate in the network. This participation not only grants several benefits with hardly any cost, but provides funding opportunities. In contrast, once funding is assured, the motivation to contribute to the network, document project activities, outputs and effects or ensure proper dissemination is reduced.
18. Stakeholders' activities. A specific series of semi-structured interviews were carried out with eight informed stakeholders in order to understand their familiarity with the NARIC network, the nature of their institutional relations with the network, their satisfaction and their expectations. Among the interviewees, one was poorly informed about the recognition centres and their network (Erasmus student network). Their low visibility to stakeholders was criticised and prompted recommendations by the evaluators.
19. Transitional arrangements. The evaluation aimed to identify 'useful lessons and recommendations for the future activities of the Network, including its ability to meet the challenges of the development of transnational education in general and of the Bologna process in particular.' After having dealt with the Bologna process, a whole chapter of the final report was devoted to transnational education. National regulation and accreditation has obviously started to react to transnational education. The analysis focuses on four major aspects of transnational education:

national treatment, transparency and market access. It points out, that the General Agreement of Trade in Services (GATS) may have a strong future impact on education. The so called Most Favoured Nation treatment clause of international law (the non-discrimination principle) guarantees equal opportunities for suppliers from all WTO Members. However, it does not require any degree of market openness.

In the following chapter the theme was expanded by dealing with the phenomenon of globalisation. Globalisation caused large scale diversification, internationalisation and digitalisation (the virtual campus) of higher education in the last decade and increased transnational education.

20. Mainstreaming provisions (see Section 2.5.5).
21. Complementary activities. The study points out that the legal framework complicates complementary activities. Under the principle of subsidiarity, there are no Community provisions imposing mutual recognition of diplomas (except for certain regulated occupations that are referred to below under the heading of recognition for professional purposes ⁽²¹⁾). For this reason there are currently no diplomas that are recognised at European level for academic purposes ⁽²²⁾. Universities, which are autonomous institutions, are entirely responsible for the content of their curricula and for awarding diplomas and certificates to students. The authorities of the Member State concerned recognise the diplomas and certificates. 'A major strength of the (Bologna) process is its complementarity with other developments in progress. It reinforces and it is being reinforced by other

⁽²¹⁾ Regulated professions: the sectoral system (nurses responsible for general care, dental practitioners, veterinary surgeons, midwives, architects, pharmacists or doctors). The general system where one of two Directives will be applicable, depends on the level of studies recognised by the diploma: either Directive 89/48/EEC for the recognition of higher-education diplomas awarded on completion of professional education and training of at least three years duration (A-levels or equivalent such as the Bac+ three years), or Directive 92/51/EEC which covers diplomas, certificates and other vocational training titles at a lower level than those covered by Directive 89/48/EEC. The contact points for the regulated professions are listed at http://europa.eu.int/comm/internal_market/en/qualifications/contact.htm [cited 3.12.2003] and in the many cases are also the NARIC centres. Lifelong learning or continuing professional development remains a matter for Member States. There have been a number of legal disputes in this field from 1997 onwards. See for example the cases of hospital administrators and pharmacists in France; use of University degrees in Germany; and Ski Instructors in Italy. http://europa.eu.int/comm/internal_market/en/qualifications/01-186.htm#3 [cited 3.12.2003].

⁽²²⁾ Even degrees from perhaps the most 'European' Higher Education Institution, the European University Institute (EUI, set up by all Member States) in Florence, Italy, are recognised in the Member States via State Treaties between the EUI and each Member State.

tools/factors which point in the same direction: Lisbon Convention, Diploma Supplement, ENQA, EU Directives, EU mobility programmes including ECTS, ENIC/NARIC network and reforms entailed by the accession process to the EU in the countries concerned.’⁽²³⁾ (Haug and Tauch; 2001).

After the programme had reached maturity, been completed or was terminated

22. Programme context (for the policy context see point 6). The evaluators scrutinised the reports of all the projects financed over the period and their quality varies considerably, spanning from excellent to unacceptable. Several recognition centres seem to be unconcerned about the quality of their reports and the feeling is gained that reporting is being undertaken as an unavoidable bureaucratic exercise. There is evidently a cultural divide here, which might be overcome by setting clear reporting guidelines.
23. Programme timing. The period observed was from 1997 to 2001 only. The activities of the evaluators started on 1 December 2001; on 1 March 2002 an interim report was delivered to the Commission and the comments received are integrated/addressed in the present final report which was delivered on 10 July 2002 in its draft form for comment.
24. Programme budget. A specific set of questions aimed to trace a profile of financing and costs of recognition centres. With the exception of the very small recognition centres, (where salaries have a lower impact on costs than in the majority of centres, while travel and conference costs have a higher impact) the size of NARIC is not directly related to the impact on the overall budget of different cost items. ‘Salaries’ is by far the item having the highest impact on the budget of recognition centres (an average 64 %). ‘Travel’ and ‘overheads’ are the second *ex-aequo* categories having the highest impact on NARIC costs, with an average 11 % each. ‘Overheads’, totalling on average 11 % of the costs of recognition centres, have a very different impact on the different centres.
25. EU priorities. The present situation is described under VET is the same as in professional recognition with regard to number of active recognition centres: 14 are active in the domain⁽²⁴⁾ (52 %) while 13 are not (48 %). The identity of recognition centres varies (i.e. there is no direct correspondence between the activity on the two domains) and perspectives of development are totally different. Recognition of qualification/degree as a guiding assessment is by far the more diffused way (75 %) to deliver assessments. In separate cross tables the monthly contacts of recognition centres amongst each other are listed.
26. Programme spill-over. There is a general move to expand the scope of the Network’s activity action beyond the narrow mandate assigned by the Socrates II decision. This development meets with the wishes of many involved but requires formal processes leading to the adoption of comprehensive terms of reference for the network, the identification of those taking part, the delegation of clear roles and responsibilities and the approval of internal rules. Structured multi-yearly planning cycles also need to be adopted. In the absence of such processes, the enlargement of the scope of the network is destined to be a source of further problems.
27. Monitoring tools. A set of multiple methods and tools had been used to monitor the networking process. For example, for ENIC the monitoring of messages exchanged over the list by the recognition centres appeared relevant to the understanding of some important points of the evaluative mandate and was therefore carried out. For more details about the methodology of the study see Section 2.5.3.
28. Programme flexibility. No *ex-ante* evaluation has been done.

⁽²³⁾ This view contrasts with ‘In order better to understand the Community rules on the recognition of diplomas, a distinction must first be made between recognition for academic purposes (i.e. you would like your title to be recognised because you wish to continue your studies) and recognition for professional purposes (i.e. you would like your title to be recognised because you wish to work in a certain profession).’ http://www.aubg.bg/_files/cep/Erasmus/Recognition_diplomas.doc [cited 7.1.2004].

⁽²⁴⁾ The text of the question was: ‘does your centre deal with recognition of vocational education and training?’

29. Systemic impact. Being part of the higher education system and in some countries situated inside Ministries, recognition centres have high systemic impact. The closer they are to government, the higher they value their impact (and the more restricted they are in undertaking initiatives not entrusted by the government. Recognition centres were requested to break down monthly contacts per area of activity per category of user. The following 10 categories of users were considered: national students; foreign students; academic staff of national Universities; academic staff of foreign Universities; administrative staff of national Universities; administrative staff of foreign Universities; organisations; career guidance institutions; other recognition centres; European Commission. Recognition centres were left free to include extra categories and some of them did so. The following additional users were included: national ministries and institutions; private enterprises, insurance companies, etc.; basic and secondary schools/education institutions; national students studying abroad; persons with a foreign qualification; social organisations; employers. The outcome was as follows:

Clients of recognition centres are mainly foreign students (43 %), followed by national students (20 %). Together these two categories make up the majority of the recognition centres' customers. A second group, 'Universities' account for 12 % of the total, which increases to 17 % if considering academic recognition only. Inter-NARIC requests for information/documentation are frequent (6 %) particularly when considering that these are only one of several ways for recognition centres to collaborate and interchange information.

30. Dissemination phasing. Dissemination of project results is not the main concern of recognition centres. A minority disseminate their reports to other centres and some confuse dissemination of results with the delivery of the activity report to DG EAC. About the 45 % of questionnaire respondents admitted to either only sending the report to EAC or to not undertaking any dissemination activity. Some of them complain of the absence of feedback from DG EAC and/or

other colleagues when sending to the network. For the purpose of the present analysis, the sending of reports to DG EAC is not considered a dissemination activity. A further 40 % of respondents say that they usually send out their reports to all of the recognition centres or as in one case they publish their reports on the NARIC website. About 15 % of respondents deliver oral presentation of the results of their activities at yearly meetings.

The evaluation process

31. Development process. Carried out with the strong involvement of the Steering Group, the NARICs themselves and stakeholders and clients.

32. Evaluation culture. Among those recognition centres subject to evaluation the majority (66 %) considered evaluation as a confidential process the results of which are not to be disclosed to third parties. Only three out of the nine recognition centres subject to a formal process of evaluation agreed to disclose their reports to the evaluators of the present exercise. Individual recognition centres were also irritated by the request for interviews and collaborated only on a personal basis. The claim to confidentiality indicates a transparent evaluation culture which has not been internalised by the recognition centres.

33. Evaluation method. The multi-method mix makes the overall picture of the evaluation more reliable. However, there were weaknesses: 10 visits have been paid to as many recognition centres. Each visit lasted about half a day which might be not have been sufficient for thorough consulting work. The semi-structured approach of interviewing different national centres yielded in-depth results for each centre at the expense of greater comparability. However, the report is comprehensive and detailed.

34. Evaluation completeness. The total list of NARIC national key points is far longer than the number of centres evaluated with a total of 30 national key points. Among the 10 countries visited there were two candidate countries (Bulgaria and the Czech Republic) and one EFTA country, Norway.

All components of the programme were evaluated: historical background, financing,

external developments, academic, professional and VET recognition. The case studies represent the major vehicle for in-depth project knowledge.

However, the length of the evaluation report (around 500 pages including annexes with case studies and project cards) raises questions about its ability to be assimilated and thus its impact.

35. Evaluation indicators. A criticism is that the term indicator was used first indiscriminately to refer to the tool or reference used for the measurement of observed elements and for the result of the measurement and observation.
36. Evaluation skills. The methodology was appropriate to the objective of the evaluation. The amount of secondary data taken into consideration is high. Questionnaires were carefully drawn up in collaboration with the steering group and covered all relevant fields. However, the objectivity of the report can be

questioned for some of the assumptions. Few passages contain value judgements, which are of a more subjective nature. For detailed information about the methodology used, see 'Methodology' in the textual analysis (Section 2.5.).

37. Evaluation censorship. No evidence given. In contrast, the high involvement and substantial comments of the Evaluation Steering Group suggests that the content was taken seriously and was disseminated.
38. Evaluation budget. An overall cost of EUR 44 915 was envisaged for the evaluation study, which is about 1.5 % of the project sum evaluated.
39. Evaluation dissemination. It is envisaged that the evaluation report is made available to the project partners and to all Member States.
40. Reporting formats. There is no evidence that the dissemination materials are being targeted to different audiences. The terms of reference require only a report.

References

- AMU Gruppen. *Evaluation of Phare VET programme BG 95.06, vocational education and training; education; research, science and technology (VETERST) in Bulgaria*, Stockholm, 15 February 2000.
- Birks Sinclair. *Strategic review of vocational education and training: Czech and Slovak Republics*. Durham: Birks Sinclair and Associates, 1993 (Final report).
- Birzea, C. et al. *Impact of the Leonardo da Vinci programme on vocational training in Romania 1997-1998*. Bucharest: National Centre of the Vocational Training programme Leonardo da Vinci, 1999.
- Decision No 1994/819/EC of the Council of 6 December 1994 establishing an action programme for the implementation of a European Community vocational training policy. *Official Journal of the European Communities*, 29 December 1994, L 340, p. 8-24.
- ETF – European Training Foundation. *Review and lessons learned of Phare VET reform programme 1993-1998*. Turin: ETF, 2001 (cross-country report).
- European Commission, DG for Education and Culture. *Valorisation strategy for the Leonardo da Vinci vocational training programme*. Brussels: European Commission, 2002 (B/3DS/ds D(2002)1484).
- European Commission. *Building of the Leonardo da Vinci programme – analysis of the programme's impact on vocational training at European and national levels*. European Commission, 2001.
- European Commission. *Final report on the implementation of the first phase of the Community action programme Leonardo da Vinci (1995-1999)*. Luxembourg: Office for Official Publications of the European Communities, 2000 (Documents COM, (2000) 863 final).
- European Commission. *PHARE – manual for contract and procurement rules, under the decentralised implementation system*. Brussels: European Commission, 1994.
- Handley et al. *Study on the opening of the Leonardo da Vinci programme to the countries of central and eastern Europe*. London: NCVQ – National Council for Vocational Qualification, 1996.
- Haug, G.; Tauch, C. *Trends in learning structures in higher education (II)*. Follow-up report prepared for the Salamanca and Prague conferences of March and May 2001.
- Ministry of Education and Science. *Bulgarian vocational education – with European standards*. Folder with information about programme BG 95.06. Sofia: Ministry of Education and Science, 1999.
- Ministry of Education, Youth and Sports. *National programme of development of education*. Prague: Ministry of Education, Youth and Sports, 2001.
- 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.
- National Observatory/NTF. *Ex-ante evaluation of EQUAL European Community initiative in the Czech Republic*. Prague: National Observatory – National Training Fund (NTF), 2001.
- NCU/NTF – National Coordination Unit/National Training Fund. *Evaluation of the first phase of the LdV programme in the Czech Republic (1996-1999) and valorisation of its results*. Prague: NCU/NTF, 2002.
- NCU/NTF – National Coordination Unit/National Training Fund. *Report on the LdV programme: Czech Republic*. Prague: NCU/NTF, 1999.
- NIE. *Scientific theoretical quarterly, strategies for policy in science and education. Summaries*. Sofia: National Institute for Education (NIE), 1999.
- NTF. *Human resource development strategy*. Prague: NTF – National Training Fund, 2000.
- OMAS. *Annual assessment report of the European Union Phare programme for assistance funded under BG 95.06, VET, education reform, science and technology*. Sofia: OMAS Consortium Southern Unit, 1998.
- OMAS. *Final assessment report for assistance funded under Phare VET reform programme in the Czech Republic*. Bratislava: OMAS Consortium – Middle Unit, 1999.
- Phare programme SR 94.03, VET reform in the Slovak Republic, ex-post evaluation, *Final report of workshop*. Bratislava: 1999.

Quarterly Support Centre for Higher Education and Information. *Public report of the Phare project BG 95.06-05*. Sofia: 1998.

Questionnaire for final evaluation, Phare programme BG 95.06. Programme valorisation meeting in Sofia, 2002.

Suurla, R.; Markkula, M. *Effective dissemination. A guide of the results on international educational projects*. Helsinki: Finnish National Coordination Unit, 1999.

Technické úkoly v souvislosti s realizací programu Leonardo na národní úrovni. Funkce přejímané Národními koordinačními jednotkami.

United Nations Development Programme. *National human development report, Slovakia*, 1998.

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: Office for Official Publications of the European Communities, 2004 (Cedefop Reference series, 3040).



Quasi-market reforms in employment and training services: first experiences and evaluation results

Ludo Struyven, Geert Steurs ⁽¹⁾

Abstract

In this paper we focus on quasi-market reforms for employment and training services for job seekers. Open tendering is the standard form for the introduction of market forces in several countries, with a competitive bidding process used to determine with which provider an order is placed for a particular service. It is an alternative to the more traditional method of subsidising providers, where a country's own government department/service, in a protected legal position, receives a guaranteed share of government funding. The focus on quasi-market reforms means that we will not cover the broader evaluation issues of active labour-market and training policies, in so far as they are disconnected from the mode of service delivery. Quasi-market changes are only just being put into place but this does not mean that any kind of evaluation is impossible. We start from a theory of quasi-markets as developed by Le Grand and Bartlett (1993) which specifies the conditions quasi-markets have to meet if they are to succeed. In our analysis, we will make a preliminary assessment of the extent to which those conditions appear to be met in practice.

⁽¹⁾ Ludo Struyven is working as a research manager in the Education and Labour Market Section at the Higher Institute of Labour Studies, Catholic University of Leuven, Van Evenstraat 2E, B-3000 Leuven; e-mail: ludo.struyven@hiva.kuleuven.ac.be.
Dr. Geert Steurs is working as a project manager at IDEA Consult, Congressstraat 37-41, bus 3, B-1000 Brussels; e-mail: geert.steurs@ideaconsult.be.

Table of contents

1. Executive summary	223
2. Introduction	225
3. The role of quasi-markets	226
3.1. The economic rationale for competitive service delivery	226
3.2. Definition of quasi-markets	226
3.3. Quasi-markets as allocation model of service delivery	226
3.4. Quasi-markets: do they work?	228
3.5. Linking contestability and market forces to the activation of benefit claimants	229
4. Overview of reforms	230
4.1. The four countries at a glance	230
4.2. Objectives of the reforms	231
4.2.1. Outcomes and costs	231
4.2.2. Choice	231
4.2.3. Other motives	232
4.3. Market forces in the broader institutional and policy context	232
4.3.1. The Swedish context of the 1980s	232
4.3.2. Australia, the Netherlands and the UK: institutional reforms in the second half of the 1990s	232
4.3.2.1. Relationship between functions	232
4.3.2.2. Scope for market forces in implementing functions	233
4.3.2.3. Market forces and the concept of the welfare state	233
4.3.3. Conclusion	234
5. Evaluation framework for quasi-market mechanisms	236
5.1. Criteria for summative evaluation	236
5.1.1. Efficiency	237
5.1.2. Responsiveness	237
5.1.3. Quality	237
5.1.4. Choice	237
5.1.5. Equity	238
5.2. Criteria for formative evaluation	238
5.3. Available evaluation studies	240
5.3.1. Australia	240
5.3.2. The Netherlands	241
5.4. Summary	242
6. Formative evaluation	243
6.1. Market structure	243
6.1.1. Separation between purchaser and provider	243
6.1.2. Provider competition	243
6.1.2.1. The relationship between tendering procedures and provider competition	243
6.1.2.2. Price versus quality competition	244

6.1.3.	Player's market	245
6.1.3.1.	Concentration and increases in scale	246
6.1.3.2.	Difficult entry to the market	247
6.1.3.3.	The position of the former public employment service	248
6.1.4.	The position of the purchaser	248
6.1.4.1.	Degree of monopoly at the purchaser side	248
6.1.4.2.	Degree of centralisation at the purchaser side	249
6.2.	Information and transparency	250
6.2.1.	Lack of transparency	250
6.2.2.	Contractual compliance monitoring in Job Network	250
6.2.3.	Performance measurement in Job Network	251
6.3.	Transaction costs	252
6.4.	Motivation: the combination of input and output financing	253
6.4.1.	Input and output financing	253
6.4.2.	The Dutch payment system	253
6.4.3.	Output financing and training	255
6.4.4.	Impact of the outcome-based focus on provider behaviour	255
6.5.	Creaming and parking	256
7.	Summative evaluation	257
7.1.	Efficiency	257
7.2.	Responsiveness and choice	258
7.2.1.	Application of the job seeker classification instrument	258
7.2.2.	Choice of provider	259
7.2.3.	Client satisfaction	259
7.3.	Quality	259
7.3.1.	Services of the Australian Job Network	260
7.3.2.	Training within Job Network	260
7.3.2.1.	The amount of training	260
7.3.2.2.	Outcome based payment and incentives to provide training	261
7.4.	Net impact and equity of the Australian Job Network	261
7.4.1.	Gross outcomes	261
7.4.2.	Sustainability of outcomes	262
7.4.3.	Before-and-after comparisons	263
7.4.4.	Provider comparisons	263
7.4.5.	Net impact and equity	263
8.	Conclusions	265
8.1.	Assessment of evaluation strategy	265
8.2.	Main findings	265
8.3.	Lessons for other countries	268
	List of abbreviations	269
	References	270

List of tables and figures

Tables

Table 1: The key functions in employment placement and spread among government organisations (situation in 2002)	233
Table 2: Objectives, instruments and players in market forces in four countries	234
Table 3: Criteria for summative evaluation of quasi-market reforms	237
Table 4: Conditions and criteria for the functioning of quasi-markets for social policy	239
Table 5: Tendering procedure	244
Table 6: Current situation	245
Table 7: C4 concentration ratio	247
Table 8: Success rates of established versus new providers (Job Network 1)	247
Table 9: Market types based on the number of providers and purchasers	249
Table 10: Payment structure and competition	254
Table 11: Cost-effectiveness of Job Network	258
Table 12: Gross outcomes of Job Network	262
Table 13: Empirical evidence concerning the impact of quasi-markets in Australia (1998-02) and the Netherlands (2000-02)	265
Table 14: Empirical evidence concerning the conditions for quasi-markets in Australia (1998-02) and the Netherlands (2000-02)	266
Table 15: Advantages and disadvantages of two market types based on tendering practice in Australia and the Netherlands	267

Figures

Figure 1: Traditional model of service allocation	227
Figure 2: Quasi-market model	227
Figure 3: Voucher model	228

1. Executive summary

In this paper we focus on quasi-market reforms for employment and training services for job seekers. Open tendering is the standard form for the introduction of market forces in several countries, with a competitive bidding process used to determine with which provider an order is placed for a particular service. It is an alternative to the more traditional method of subsidising providers, where a country's own government department/service, in a protected legal position, receives a guaranteed share of government funding. The focus on quasi-market reforms means that we will not cover the broader evaluation issues of active labour-market and training policies, in so far as they are disconnected from the mode of service delivery. Quasi-market changes are only just being put into place but this does not mean that any kind of evaluation is impossible. We start from a theory of quasi-markets as developed by Le Grand and Bartlett (1993) which specifies the conditions quasi-markets have to meet if they are to succeed. In our analysis, we will make a preliminary assessment of the extent to which those conditions appear to be met in practice.

This paper reports primarily on evaluation of the Australian quasi-market model, which is in many ways the most radical in the world. Closest to Australia are the reforms in the Netherlands. In the UK (New Deal programmes) and Sweden (training), the model has been adopted only partly and with divergent features along the employment and training programmes. A system analysis is provided of quasi-market arrangements in these countries. Subsequently, the evidence comes mainly from the Australian Job Network and, to a lesser extent, the emerging private reintegration market in the Netherlands.

The experiences from the Australian and other quasi-market reforms support the view that the market model for providers of reintegration services is superior to that operating previously. However, there is no hard empirical evidence for this. It is extremely difficult to distinguish rigorously between the effect of market forces and other changes that may have taken place simultaneously, such as changes in the benefit system

or in the labour market. But this can not be a justification for the lack of external (scientific) evaluation research on these reforms.

In Australia, the most significant gain is that the market system leads to more cost-effective performance: services cost less, with a more or less similar net impact as previous programmes and sufficient responsiveness and quality. Weaker points are the very limited choice for job seekers and the erosion of the level of services, especially with regard to job-specific training. However, findings on net impact and quality are tentative and should be added to other findings on the amount of training services, the financial incentives and their role for disadvantaged job seekers. It was found that financial incentives, focusing on employment outcomes, have an adverse effect on the equal treatment of job seekers. The incentive structure favours the pursuit of short-term employment outcomes over training and education, although training and education may, in the longer-term, be of greater benefit to job seekers.

In the Netherlands, it is not yet clear how far the arrangements to foster responsiveness and choice between providers may work well. The Dutch reintegration market tends to become self-regulating on quality: at the moment, there is no evidence of increased efficiency nor equity.

Despite these mixed findings, there is more positive evidence on the conditions required for the functioning of quasi-markets. Typical problems concerning information and transparency and transaction costs can be overcome by refinements in structure, tendering mechanisms, implementation and monitoring arrangements. In this respect, the experiences are not similar, as shown by the market patterns in Australia and the Netherlands. Australia has one centrally operating purchaser, in contrast to the Netherlands where there are several purchasers operating in a decentralised manner. They represent contrasting market types, with both advantages and disadvantages.

The evaluation review of emerging markets for employment and training services highlights the

central role played by incentives. Therefore, it makes sense to give much more attention in evaluation research to the relationship between

tendering and market conditions on the one hand and the behaviour of both providers and job seekers on the other.

2. Introduction

Countries are increasingly turning to private organisations to implement activation and reintegration measures aimed at job seekers. In several OECD countries, this is currently taking the form of a market forces policy, based on the creation of a market for employment placement and reintegration. The mechanism used is that of open tendering, with reliance on private bodies (partially) replacing government organisation for employment placement, training and reintegration. The central issue of this study deals with the experiences acquired with market forces. Four pioneering countries are examined: Australia, the Netherlands, Sweden and the UK. The key questions are as follows. Within which structural reforms is the introduction of market forces placed? What are the intended objectives of the market forces? Is there a genuine market for reintegration services? What about the former state provider? What are the impact of the tendering rules and the incentive structure on this provider? What are the effects of market forces on the workings of the labour market, in particular the position of the job seekers? What is the impact on the services provided? What happens to the amount and the nature of training?

Particular attention is given in this study to evaluating the implementation and (potential) impact, based on both formative and summative evaluation. Hence, we examine to what extent during implementation formative or summative evaluation has taken place and in what way this has already influenced further implementation.

This paper is structured as follows. In Chapter 3, we deal primarily with the economic rationale of quasi-markets, and summarise the evidence so far with similar reforms in related policy sectors of other human services. We then give a short overview of the main reforms and policy objectives for the quasi-market reforms, followed by a discussion about the institutional and policy context of the benefit and activation systems of the countries under consideration. In Chapter 5 we introduce framework for evaluation, based upon the theory of quasi-markets. Subsequently, we look in more detail at the empirical findings concerning formative (Chapter 6) and summative evaluation (Chapter 7). In Chapter 8, we discuss concluding points.

3. The role of quasi-markets

3.1. The economic rationale for competitive service delivery

Market forces are used to open up government organisation to the market. In this context, we understand market forces as being a means of achieving public goals not as shifting financing from the public sector to the private. Traditionally, services aimed at labour reintegration have been heavily monopolised by government provision. In a system of market forces, the services are offered by a wide field of providers, with both NGOs and profit-making organisations (e.g. employment agencies) operating alongside the government provider. Market forces do not mean that activities are 'left' to the market, but that the government has the services provided by private market parties. The government can also continue to play its own role as provider. The introduction of market forces stems from the pressure to improve efficiency and responsiveness in the public sector. In economic literature, the role of competition and contestability is threefold:

- (a) using competitive pressure to constrain activities and drive producers to keep prices down;
- (b) organising according to the best known techniques of production;
- (c) innovating to avoid economic annihilation (Webster and Harding, 2000).

Markets with a plausible threat of entry may be enough to be described as contestable markets. Government funding is justified on the grounds that purely private provision will result in sub-optimal consumption levels where positive social externalities are present and economic justice principles are relevant. Therefore, a broad distribution of education, health and social services may be required, broader than would arise under purely private arrangements. However, while public funding is justified, it does not immediately follow that the government must also 'provide' all or any of these services. In the literature this model has been called quasi-markets.

3.2. Definition of quasi-markets

In a quasi-market, the relationships between purchasers and providers are structured in a commercial way. This produces a shift from bureaucratic monopoly providers towards independent, competitive providers: 'The government bureaucracies responsible for providing these services were broken up, being replaced by state purchasers who contracted with independent providers to provide the service concerned' (Le Grand, 2001). Decisions are made on the basis of the market criteria of price, quality and delivery time. As in an ordinary market, competition exists between various providers, but these are not, however, necessarily driven by profit maximisation. On the demand side, the available budget is distributed by a purchasing agency. Le Grand and Bartlett define this new mode of public service delivery as an internal or 'quasi'-market: a market in the sense of independent agents competing with one another for custom from purchasers, but a quasi-market in that, unlike in a normal market, purchasing power comes not directly from consumers but from the state (Le Grand, 2001). Other differences from normal markets are the appointment of an agent to act as a purchaser on behalf of the final consumer and – in some policy sectors – a preponderance of non-profit or even public providers. The new delivery system took off in many countries in the 1980s and 1990s in social policy fields like health, education and housing. According to the theory of quasi-markets, market forces have two main goals: on the one hand, to increase effectiveness and, on the other hand, to increase the responsiveness of the providers and the freedom of choice for consumers (Le Grand and Bartlett, 1993).

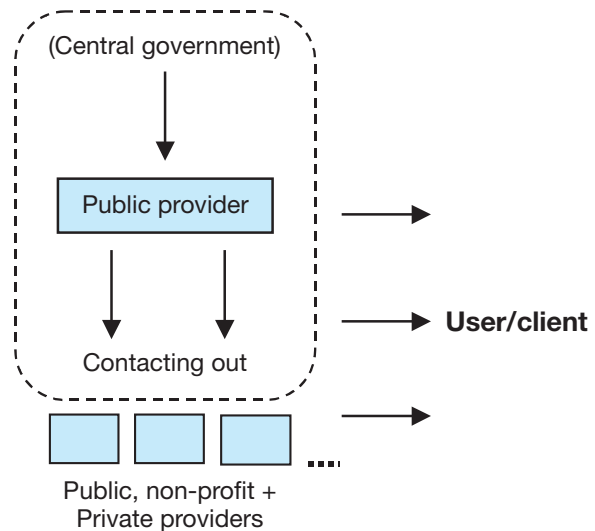
3.3. Quasi-markets as allocation model of service delivery

The introduction of a quasi-market for public funded services can be considered as a specific

mode of service delivery, to be distinguished from privatisation or contracting out. With privatisation a former public sector function is moved permanently into the private sector with a change in the owner's structure and in the way of financing. Contracting out is often understood as outsourcing from the public provider; in a broader sense contracting out means that the former public sector function is subject of private sector competition, where the public provider can continue to act as one of the competing actors. In this sense contracting out stands for a quasi-market arrangement. Finally, there is the voucher system, which enables the client or customer to shop around between providers (Derksen et al., 2000; Thuy et al., 2001; Green et al., 2001; Struyven et al., 2002). These models can be seen as different allocation models: resourcer oriented (the traditional government's provider); provider and purchaser oriented (quasi-market model) and customer oriented (voucher model).

In each model the following roles are distinguished: the resourcer role, the purchaser role, the provider role and the user role. In the traditional model, publicly funded services are provided by the public provider, with some parts usually being contracted out to other providers.

Figure 1: **Traditional model of service allocation**



In this case, the government is combining the role of resourcer, purchaser and provider. In the quasi-market model, a clear distinction is made between the purchaser (principal) and the provider (agent). In the voucher model, funds are going directly to the user, who can make a choice between different providers in the market. In this model, the user (or customer) is acting as a purchaser of services.

Figure 2: **Quasi-market model**

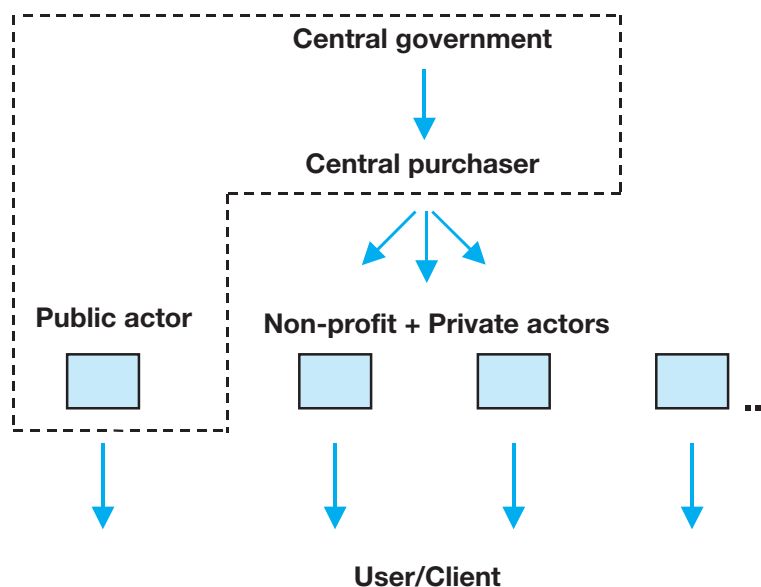
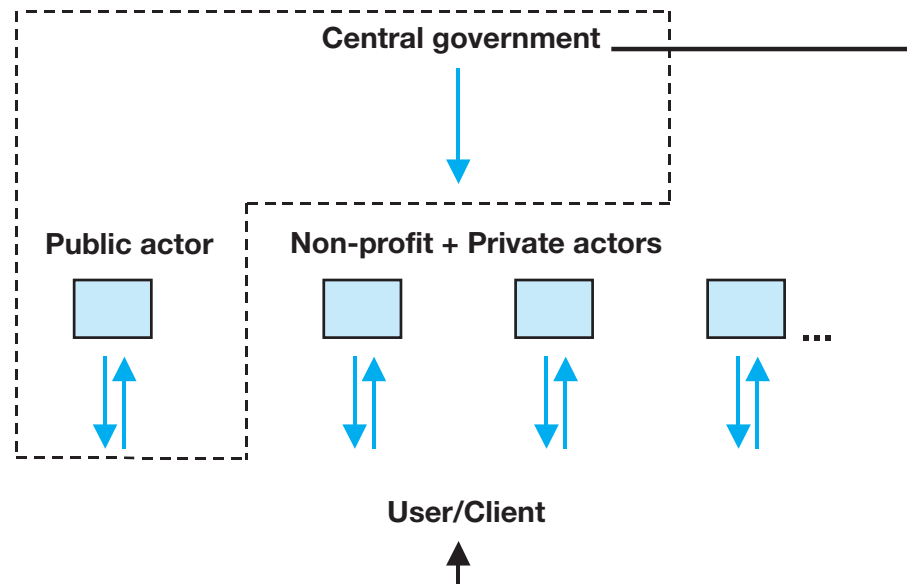


Figure 3: Voucher model



3.4. Quasi-markets: do they work?

Market forces cannot simply be equated with less social protection. What is striking is that the countries in which quasi-market arrangements are implemented, are involved in far-reaching reforms to systems of benefits, activation and employment placement (Struyven et al., 2002). The application of market forces can be seen as a means of making social security manageable: not by breaking down rights to benefits, but by reorganising implementation. Rather than 'free market forces', the issue is much more one of 'managed market forces' or even 'controlled market forces' (see the concept of managed liberalisation in van der Veen, 2000). The question is not therefore 'government or market' but of 'which policy mix of government and market?'.

Here we are mainly concerned with quasi-markets in typical publicly funded sectors, such as education, health and other human services. British experiments in these areas are among the most radical of those tried. In his review of 2001, Le Grand reports on the evidence so far with these quasi-markets. The most striking conclusion from this review is how different they are (Le Grand, 2001). In the British health care market reforms, there seems to have been little overall measurable change. The quasi-market delivered

relatively little of the benefits that its advocates hoped, but also generated few of the disasters that its critics predicted. In contrast, the market in education does seem to have brought significant changes in behaviour in British schools. There have been gains in efficiency and choice, but also significant equity losses (see below with regard to the evaluation criteria). According to Le Grand, there are a number of possible explanations for the relative failure in health care and the relative success in education. First, the incentive structure was relatively weak in health and relatively strong in education (e.g. reward for success, penalty for failure). Second, the constraints faced by the agents were different (e.g. degree of autonomy; flexibility to respond to demand pressures). Finally, the motivations of the active agents played a role. In hospitals, the dominant motivation is quality of the care provided by specialist consultants and their professional advancement. In education, school 'managers' were more preoccupied by the financial success or failure of their institution. Le Grand concludes that, for the introduction of quasi-markets into public service delivery significantly to change behaviour in the relevant actors, a number of conditions have to be fulfilled. In this paper we will examine how far these conditions are present in the emerging employment and training services markets.

3.5. Linking contestability and market forces to the activation of benefit claimants

The practice of market forces is rooted in the broader activation concept (European Commission, 1998). The shared goal of the reforms is to allow social integration to take place more via participation in the labour process than purely by redistributing income through benefits. The efforts of the job seeker, the reintegration company and the benefits agency have to be geared towards the same goal. The premise of 'work first' (UK) or 'work before income' (the Netherlands) figures in each country. Incentives or financial stimuli deeply influence the nature of

the mutual relationships between the players involved. Reintegration companies are steered towards achieving the placement standard and job seekers are judged on their willingness to accept an activation offer or a job. The outcome takes precedence. The underlying reasoning is that steering based on placement results is more easily based on competition between providers; after all, their role is ultimately that of mediator with the demand for labour. The benefits agency is given more autonomy and responsibility in exchange for result commitments, while the duties of the individual job seeker as 'beneficiary' are now also considered. Broad unanimity exists concerning the goal, i.e. linking the right to a benefit to participation in an integration route.

4. Overview of reforms

4.1. The four countries at a glance

The Australian Job Network is an international model for market forces in employment placement. The model has been in place since 1 May 1998 for all the employment placement and reintegration services. Two tendering procedures have already taken place, in 1997 (Job Network 1) and in 1999 (Job Network 2). The third round is planned for March 2003. This case study is targeted at three labour-market services which were deployed in the market. 'Job matching' consists of placement services for all registered job seekers, including vacancy recruitment and helping job seekers to find a job. 'Job search training' is intended for job seekers with a moderate degree of disadvantage and is aimed at imparting 'job search skills'. 'Intensive assistance' involves individual counselling for the most disadvantaged job seekers. Dramatic institutional reforms were carried out at the same time as the introduction of Job Network. The folding of the Commonwealth Employment Service and the formation of Centrelink, an autonomous government organisation which provides a wide range of services to citizens through the 'one-stop shop' format, formed part of this. The Department for Employment, Workplace Relations and Small Business (DEWRSB) acts as sole purchaser. Centrelink refers job seekers on to the providers: Employment National (successor to the Commonwealth Employment Service), private providers and NGOs.

The Dutch government has set itself the aim of creating a 'purely private reintegration market'. At national level, the principle of open tendering has been generalised for job seekers and the disabled (first round in 2000; second round in 2001). For those on benefits, at municipal level the evolution towards 'transparent purchase' of the reintegration services present on the market is still very gradual. The 'privatisation' of labour reintegration forms one of the pillars of the Structure of work and income implementation law (SUWI law),

which entered into force on 1 January 2002. It is hoped, through intervention in the administration organisation, to reduce the sharp increase in numbers of disabled people receiving disability benefits. The population of job seekers and the disabled is part of the responsibility of the new benefits agency (UWV or Employees' Insurance Benefits Agency); benefit recipients remain under municipal responsibility. Paying out benefits – and associated tasks such as establishing the right to benefits – remains within the public domain. Reintegration as a whole is transferred into the private domain. The reintegration component of the public employment placement organisation (*Arbeidsvoorziening*) is transferred to the autonomous public company KLIQ. Several basic tasks, such as registration and direct placement are offered jointly for both groups via the one-stop shop of the Centres for Work and Income (CWI), which are taking over some of the tasks of the former job centres. The Netherlands has many buyers as a result of shared commissioning by the UWV and the municipalities.

The Welfare to Work programme of the British Labour government to a large extent continues the trend of involving private parties in implementing the various New Deal programmes. The New Deals fall under the auspices of the Employment Services, which itself retains a number of tasks regarding intake and counselling. The Employment Service is an integrated benefits and employment placement organisation and is organised along the one-stop shop concept through a network of Jobcentres Plus. The service operates as an autonomous component of the new Department for Work and Pensions. Since 1998, the Employment Service has been organising individual tendering procedures per New Deal programme (for young people, the long-term unemployed, single parents, etc.) and per region. What exactly is carried out by the market can vary from one region to another. In addition, the Department for Work and Pensions sets up pilots for employment zones and action teams. These are initiatives for the placement and reintegration of job seekers in a well-defined area

of high structural unemployment. The benefits function is also privately organised.

Sweden has for a long time emphasised active labour-market measures, which involves devoting a great deal of attention to training. The Swedish case study concentrates on the introduction of market forces in organising the supply of training for job seekers. In contrast to the other countries, this reform did not take place within the Employment Service. It was in two phases, starting in 1986 when organisation of training was reformed for the first time, with separation of demand and supply. Publicly financed training became the competence of the National Labour Market Board and 24 County Labour Market Boards. A new agency, AmuGruppen, was formed, with the task of providing training courses. At the same time, the market for publicly financed training courses, via decentralised tendering, was opened up to public and private training providers. In 1993, the second phase of the reforms was implemented. The autonomous government agency was transformed into an independent state enterprise (now Lernia), which has to compete under the same conditions with other players in the training market.

4.2. Objectives of the reforms

In the theoretical literature about market forces, two main aims are usually identified: to increase effectiveness and to increase the responsiveness of providers and freedom of choice for consumers. These aims partly correspond with the aims of active labour-market policies in general: to improve the employability of the unemployed, increase the efficiency of job searching (in contrast with passive policy, which provides benefits for those without work), secure job outcomes and improve equity (Productivity Commission, 2002).

4.2.1. Outcomes and costs

When studying the objectives for market forces in terms of reintegration, the motive of technical efficiency and cost control seems to rise to the surface in all countries. In addition, countries also aim for greater allocation efficiency using a new service concept (Australia) or a more flexible supply of training (Sweden). Australia and the UK

both emphasise the importance of outputs and outcomes (results). The British system, moreover, is highly target-driven. In the Netherlands, market forces – via a reorganisation of the entire implementation structure – are focused on controlling the problem of disability benefits (under the WAO) which has dominated social security policy for over a decade. It is hoped that market forces will reduce those claiming benefits under the WAO (by making the employer the principal) and increase those leaving the benefits system under the WAO (by making the benefits agency the principal). Result steering and market forces go hand-in-hand in this case.

The creation of open competition is a key element in all countries. This is explicit in the policy theory of the Netherlands: the creation of a purely private reintegration market and transparency in allocating public resources function as the central objectives. Neither the Netherlands nor the UK use market forces in the context of major government budget savings. In Australia, savings as a motive for market forces were also not officially on the agenda but, in practice, significant savings do seem to have been made. The Job Network system was actually introduced by abolishing a number of existing labour-market programmes. Moreover, the government managed to bring the cost price to a little lower than it had previously been, using assumed price margins (see below).

4.2.2. Choice

Except what theory would lead us to suppose, relatively little reference is made to the second aim in the four case studies: the aim of increasing the options for the consumer. The logic is that when the government offers services as a monopolist, job seekers cannot make their own choices because there are no alternatives. The market should be better able to guarantee individual freedom of choice but little of this can be seen in present practice. It should however be said that experiments are under way with person-related or person-following budgets (e.g. personal job account in the employment zones; person-related reintegration budget in the Netherlands). Australia does refer to the goal of greater freedom of choice but, in practice, this does not seem to work well (see below). One possible explanation for the missing dimension of freedom of choice is that market forces systems

are heavily geared towards controlling reintegration supply via regulation of providers. One other explanation is that the activation context (by emphasising rights and obligations) and the job-seeking population (as a result of the presence of less independent target groups) lend themselves less to this.

4.2.3. Other motives

A further consideration concerns the operation of the government provider. In a market discussion, we could expect dissatisfaction with the government provider to be the main reason for reforms and the goal of shifting the production of labour-market services to the private sector to be a response to the problems of government provision. This does not seem to feature in official discussions in the countries in question, though some reference is made in the appeal for strong private involvement. It is likely that in practice it has been an important breeding ground for greater openness with respect to more intense market forces.

Finally, there is the political-ideological motive. Market forces are increasingly being introduced, with Australia, New Zealand, the UK and the US playing a leading role. Nonetheless, the ideological argument has to be put into perspective when looking at the prior development process in Australia or the UK. More decisive in nature is the question of how far there is a potential market for it.

4.3. Market forces in the broader institutional and policy context

The most striking aspect of the countries under consideration here is that the introduction of market forces is associated with a dramatic institutional reform of the broader system of benefits, activation and employment placement. In essence, it affects the structure of organisation and incentives. The choice for market forces is not self-evident. In this respect, we draw a distinction between Sweden and the three other countries (Struyven, et al., 2002).

4.3.1. The Swedish context of the 1980s

The context for development of market forces differs. For instance, the initial driving force

behind Swedish training policy appears to have been increasing unemployment, which created an increasing need for training that could not be met by AmuGruppen, the state provider. In addition, the Swedish authorities wanted to make the government provider more cost efficient by exposing it to competition. However, in the Swedish example, market forces seem to be contained within a process of institutional change. This sought a clearer distinction between education and training geared towards the labour market, which led to the shift in responsibility for the former Amu centres from the Department of Education to the Department of Employment.

4.3.2. Australia, the Netherlands and the UK: institutional reforms in the second half of the 1990s

These wider institutional reforms are found in the three other countries and the case studies show the extent to which Australia, the Netherlands and the UK have recently become involved in far-reaching reforms which are not yet complete. The motives for transferring to market forces are rooted in diagnosis of the existing institutions: the legitimacy of keeping certain tasks in the public domain (or not), the nature and efficiency of incentives concealed within them, the relationships between inter-related players. The structures vary in the countries studied with differences in the organisational division between functions, the scope for market forces in the implementation of functions and the new concept of the welfare state.

4.3.2.1. Relationship between functions

The relationship between the two functions of benefits and placement represents a fundamental and constantly recurring field of tension in labour-market policy. The OECD usually identifies three main functions for policy, irrespective of who is assuming them: the benefit payment function, the placement function and the referral function (to labour-market programmes) (Fay, 1997). Developments in market forces allow us to identify three additional functions:

- (a) the transparency function with respect to the vacancies market, specifically via the presence of a broad public vacancy database and/or private databases;
- (b) the tendering function: in present-day market forces, this means putting out to tender activ-

ities in the field of employment placement and reintegration;

- (c) the reintegration function: this means fulfilling and/or implementing activities in the field of reintegration (introduction, training, etc., through to placement and after-care).

Two sub-functions can be distinguished as part of the placement function: direct placement and reintegration. This means that six key functions can be identified. Table 1 examines how the functions are distributed in the new setting of organisations.

In Australia and the UK, powers are distributed between the central department and one integrated government organisation: Centrelink (Australia) and the Employment Service (UK) with the network of Jobcentres Plus. This development contrasts with the new structure in the Netherlands, where the benefits and placement functions are (remain) divided between two different organisations: the UWV and the CWI. The special feature of the Dutch situation is that two other organisational mergers are taking place: first at the level of the benefits agencies and, second, at the level of the intake and referral function. In the new configuration, the intake and referral function is the job of the CWI, whereas these tasks used to be exercised in a fragmented manner by the employment agencies, the administration agencies and the municipal social services.

4.3.2.2. *Scope for market forces in implementing functions*

Differences exist in the extent to which functions are left to the market. In this respect, Australia goes the furthest, with typical tasks related to employment placement by a public institution reduced to a minimum. In the Netherlands, intake encompasses a broader procedure and the benefits agency can also continue to carry out case management (client follow-up right through to placement). In the UK too, the role of the Employment Service remains more significant; though this varies depending on the region. However, the UK is also experimenting on a territorial basis with tendering out both reintegration and benefit tasks (Employment Zones).

4.3.2.3. *Market forces and the concept of the welfare state*

Other differences emerge from the broader perspective of social protection. It is clear that the reforms in Australia push the mutual connection between result-orientation, free competition and conditional benefits (*mutual obligation*) to the forefront. In the Netherlands and the UK, official policy discussions are moving in the same direction, although it cannot be said that a policy of less social protection is being pursued. Rather, we are seeing strong emphasis on the mutual rights and obligations of activation bodies and job seekers.

Table 1: **The key functions in employment placement and spread among government organisations (situation in 2002)**

	Australia	The Netherlands	United Kingdom
Benefit	Centrelink	– UWV – Municipalities	Employment Service (ES)
Direct placement	By tender	CWI	Jobcentre Plus (= ES)
Intake and referral	Centrelink	CWI	Jobcentre Plus (= ES)
Vacancy market transparency function	Central department	– Private market – CWI	ES
Tendering	Central department	– UWV – Municipalities	– ES – Central department
Reintegration	By tender	By tender	By tender

4.3.3. Conclusion

The introduction of the market mechanism does not stand alone, but is part of the search for a more coherent benefits and activation system. This system, in turn, is related to changes in the concept of the welfare state. Institutional reform represents one of the major pre-conditions for market forces. This, however, complicates evaluation of the impact of the market forces mechanism because the introduction of the market mechanism is difficult to isolate from the other system changes.

Table 2 presents a brief summary of the four case studies in terms of the following characteristics: objectives, type of market mechanism, commencement date, principal, frequency, purchaser/provider separation, type of market forces, definition of the assignment, target groups in the specification, quality monitoring body, role of the social partners, referral body, choice for the provider, choice for the customer and vacancies. Most of these points will be discussed in more detail later in the paper.

Table 2: **Objectives, instruments and players in market forces in four countries**

	Australia	The Netherlands	United Kingdom	Sweden
Objectives	– Geared towards outcomes – Efficiency/savings – Individual services of high quality	– Purely private reintegration market – Disability benefits cost control – Transparency in allocation of resources	– Target-driven – Geared towards result – Cost control	– Cost control/savings – Flexibilisation of training supply
Market mechanism	Open tendering	Open tendering (in UWV; in municipalities still exceptionally)	Open tendering (in theory)	4 types: Open (usually), restricted, negotiated, selective
Introduction of open tendering	Generalised since 1997	Generalised since 1999 (UWV) Municipalities: still being developed	Generalised since 1998	Generalised since 1993
Principal	One procedure through central department (Commonwealth)	– Disabled (Occupational Disability Insurance Act) + unemployed (Unemployment Insurance Act): through benefits agency (UWV) (Central ZBO – Autonomous Managing Body) – Benefit claimants: municipalities	– New Deal: Through regions of Employment Service within uniform procedure – Employment Zones + Action Teams: central department	– Regions: by County Labour Boards within uniform framework – Additional: Local Employment Services
Frequency	One national tender every three years	– one tender annually (UWV) – municipal tenders at any time	Tenders per programme every three years	Tenders per region annually
Purchaser/provider distinction	Yes	– UWV: yes – Municipalities: some still in role of provider (municipal companies)	Yes	Yes

Type of market	Monopsony, buyers' market	Towards quasi-market	Intermediate form	Monopsony per region
Definition of assignment	Five separate services: job matching; job search training; intensive assistance; new enterprise incentive scheme; project contracting (harvest labour)	– UWV: one reintegration service in five stages: reintegration plan, diagnosis; supply reinforcement, mediation; placement and after-care – Municipalities: variously specified	Specified	Specified by training
Target groups in specification	Not specified	22 target groups (Occupational Disability Insurance Act/Unemployment Insurance Act)	Separate plans per New Deal target group: young people; 25+; single parents; partners of the unemployed; disabled people; 50+	No target groups specified
Quality monitoring body	Principal	– Principal – Branch organisation – Advisory council (Work and Income Council)	– Principal – Adult Learning Inspectorate	– Principal – National Labour Market Board
Social partners	No structured advice, but bilateral consultation	Until 2001: relative involvement of sectoral councils From 2002: advisory council (RWI - Work and Income Council)	New Deal Task Force	Structured advice at national and regional levels
Customer referrals	Centrelink	UWV (in practice still by five benefits agencies) + municipalities	Jobcentre Plus	Employment agency
Compulsory participation by provider	Yes	No	Yes	Yes
Choice for customer between providers	In theory yes, limited in reality	If possible, choice between several providers	Choice of options	No choice
Vacancies	Compulsory use of registered vacancies in national vacancy database; provider brings in most vacancies	Free vacancy recruitment Several databases (private)	Free vacancy recruitment	–

5. Evaluation framework for quasi-market mechanisms

Our focus on quasi-market reforms for employment and training services for job seekers means we will not cover broader evaluation issues of active labour-market and training policies, in so far as they are disconnected from the mode of service delivery. Quasi-market changes are only just being put into place but this does not mean that any kind of evaluation is impossible. We start from a theory of quasi-markets as developed by Le Grand and Bartlett (1993) which specifies the conditions quasi-markets have to meet if they are to succeed. In our analysis, we will make a preliminary assessment of the extent to which those conditions appear to be met in practice.

This paper reports primarily on evaluation of the Australian quasi-market model and, to a lesser extent, the Netherlands. There are several reasons for this. In the first place, the Australian Job Network is in many ways the most radical in the world. Closest to Australia are the reforms in the Netherlands. In the UK, the model has been adopted only partly and with divergent features along the employment and training programmes. Also, the models in both the Netherlands and the UK have only recently undergone a similar reform. As a consequence, and in contrast to in Australia, there is no systematic review of the evidence relating to the quasi-market in the Netherlands or the UK; existing evaluation studies did not suit a meta-analysis, especially on impact evaluation. For these reasons we will focus mainly on the evaluation of Job Network; for the other experiences we rely on our own overview of the four countries as far as this is useful for the implementation evaluation (Struyven et al., 2002). For Australia and the Netherlands, an attempt has been made to gather as much reliable evaluation evidence as possible from different sources. The last paragraph of this chapter summarises the main evaluations available for both Australia and the Netherlands.

5.1. Criteria for summative evaluation

The first question on evaluation is what might be meant by 'success', or, what are the appropriate criteria for the evaluation of quasi-market reforms? A further distinction can be made between different types of evaluation, such as formative and summative evaluation. According to evaluation literature (Patton, 1982; Clarke, 1999), formative evaluations aim to providing information for programme improvement, modification and management. The question is 'How can the programme be improved?', whereas the question for summative evaluation is 'Should the programme be continued, and if so, at what level?'. Summative evaluations (or impact evaluations) aim to determine programme results and effects, especially for the purpose of making major decisions about programme continuation, expansion, reduction and funding. These two types of evaluation are not mutually exclusive. Over time, a particular programme or reform might be involved in activities from different types of evaluations, starting from procurement and probity audit, monitoring and formative evaluation to summative evaluation and meta-evaluation.

The aim of quasi-markets can be both efficiency and responsiveness, and freedom of choice and equality: 'The aim was to improve efficiency, choice and responsiveness, and to raise quality standards – and without any adverse consequences, especially for equity' (Le Grand and Bartlett, 1993). This provides four criteria for evaluation and a more recent paper by Le Grand (2001), added quality as a fifth. The corresponding five criteria can be understood as criteria for summative evaluation, stated in Table 3.

Each of the criteria is further explained and illustrated by one or more key indicators, without making any claim to be exhaustive.

Table 3: **Criteria for summative evaluation of quasi-market reforms**

Criterion	Indicator	Aim
Efficiency	Cost of service provision Value for money	To minimise total costs To minimise costs of providing a given quality or quantity
Responsiveness	Individual needs and preferences Local needs	To make the service more flexible and sensitive to needs
Quality (process, output, outcome)	Kind of services Intensity of services Client satisfaction Complaints procedures Gross placement ratio Net placement ratio Drop-out ratio	To enhance both process and output of the service
Choice	Choice of service Choice of provider for clients Choice of clients for providers	As an end in itself As an instrument for achieving the other aims
Equity	Concentration on those with the greatest needs	Increased efficiency, responsiveness and choice without adverse consequences in terms of increased inequity To avoid creaming
Source: Le Grand and Bartlett (1993); Le Grand (2001)		

5.1.1. Efficiency

Efficiency can be interpreted in many ways, but Le Grand and Bartlett distinguish two basic concepts that underlie most of these different interpretations: the total costs of service delivery and the costs of providing any given quantity or quality of a service (productive or technical efficiency). One can further distinguish between different forms of technical efficiency. Technical efficiency is achieved when a given output is produced using minimal deployment of input (input efficiency) or when maximum output is achieved using a given quantity of input (output efficiency).

5.1.2. Responsiveness

Increasing responsiveness can be regarded as the pursuit of a greater allocative efficiency. Allocative efficiency refers to the output which best fits in with the needs and preferences of the end users as well as the locality in which they operate.

5.1.3. Quality

Contrary to Le Grand and Bartlett's theory, quality cannot be left out when discussing the evaluation of public service delivery. This seems to be recognised in Le Grand's recent overview of evaluation findings of quasi-markets (Le Grand, 2001). Quality is a broad concept with many interpretations. One of the many applications, as used by the European Foundation for Quality Management, covers process and organisation criteria (enablers) as well as results criteria. For the evaluation of quasi-markets, quality indicators can be based on the service (kind and intensity of service), the client's perception (client satisfaction, complaints procedures) and (micro)-economic impact (gross and net placement ratios, drop-out ratio).

5.1.4. Choice

The logic is that when the government offers services as a monopolist, clients cannot make their own choices because there are no alternatives. The market should be better able to guarantee individual freedom of choice. After all, the

market mechanism is sustained by differences in individual preferences. As stated by Le Grand and Bartlett, a concern for choice may be justified as an end in itself, or as an instrument for achieving other policy aims. Theoretically, one can distinguish between the choice of service, the choice of providers (for clients) as well as the choice (of clients) for providers.

5.1.5. Equity

Equity rarely appears explicitly as one of the policy objectives of the quasi-market reforms, as Le Grand and Bartlett noticed: 'no doubt because they are being driven by a government for which equity considerations are not necessarily a high priority' (Le Grand and Bartlett, 1993). The key concern when providing social services is to concentrate on those with the greatest needs. Le Grand and Bartlett define an equitable service as one whose use is determined primarily by need and not by irrelevant factors such as income, socioeconomic status, gender or ethnic origin. The more care resources an individual requires to bring his or her level of welfare up to some predetermined level, the greater is his or her need. The aim is to avoid creaming and parking (see below) of the easy-to-place clients.

5.2. Criteria for formative evaluation

Another set of criteria, from a formative evaluation viewpoint, can be derived from the conditions for the functioning of markets. According to Le Grand and Bartlett, the following five conditions must be satisfied in order to be able to function as a market: there must be a market structure based on competition and prices, with accurate and complete information, with moderate transaction costs, a balanced motivational structure and without adverse consequences in terms of creaming. Theoretically, a number of problems can arise in these markets. These problems relate to the market structure, information and transparency, transaction costs and risk selection.

The main problems relate to competition and price as the basic components of a market structure:

- (a) dominant providers exploit their monopoly power due to a lack of competition among

providers, or else large purchasers exploit their monopoly power due to a lack of competition among purchasers.

- (b) the relationship between the two dominant players on both sides can be too intimate, so that the system could become one simply of decentralised budgets, with a management contract between purchaser and provider, but with no real competition on either side;
- (c) providers are driven out of business, the system fails to attract new entrants and prevent the 'exit' of efficient providers because a dominant purchaser may sour relationships with providers and lower their morale and motivation;
- (d) prices are not formed directly between producers and consumers, so do not reflect user preferences, and are 'administered' or negotiated between purchasers and providers;
- (e) providers do not necessarily face hard budget constraints because of their specific ownership status.

Besides the problems of the market structure, providers may look for strategies to reduce uncertainty or competition, for example through mergers and collaboration. These strategies can occur when it is difficult for purchasers to monitor quality and to carry out checks because of incomplete information. Providers may engage in what Williamson (1975, 1985) calls opportunistic behaviour, exploiting their informational advantage to reduce costs at the expense of quality. Following Le Grand and Bartlett (1993), there are two kinds of opportunistic behaviour to which the theoretical literature draws attention: moral hazard and adverse selection. Moral hazard occurs where providers put fewer resources into the provision of the service than is consistent with the terms of their contract. Because contracts are incomplete, i.e. they do not prescribe in full detail what providers must do under all possible circumstances, providers can reduce their costs by cutting their services within the limits of the contract. Employment service providers for instance may be obliged by contract to have regular contact with job seekers, but they may be inclined to limit these contacts to a minimum to reduce their costs (contact over the phone instead of face-to-face contact). Adverse selection occurs where providers possess certain

characteristics that may adversely affect the provision of the service and that are known to them but that they do not reveal to the purchaser. Because of this asymmetric information, the contracting conditions offered by the purchaser will target the 'average quality provider'. Providers which can offer better quality, though at a higher price, will not be interested while for providers offering poorer quality, the contract is attractive. As a consequence, the purchaser contracts the providers he does not want.

In theory both problems could be prevented in the formulation of the contract and by continuous monitoring, but this will increase administration and transaction costs. Fully specified cost-per-case contracts would then replace the incomplete block contracts. The transactions which take place in quasi-markets are often quite complex and multi-dimensional, given the complexity of the social services. Transaction costs may be particularly acute in the presence of uncertainty and risk. Both purchasers and providers, which operate in an unpredictable context of labour demand and supply, are unable to specify all future contingencies in a written contract. This problem is known as the incomplete contract: contracts are deliberately left

incomplete and reflect a willingness to cooperate rather than the precise allocation of risks, responsibilities and rewards (similar to the implicit contract under the pre-market system) (Walch et al., 1997).

Quasi-markets may also be subject to the problem of risk selection: the provider's selection of the most profitable clients. This problem, also known as cream-skimming, discriminates against the more expensive users: the long-term unemployed, the low-educated, minority groups, etc. The predominance of users' interests in providers' financial or organisational interests can also occur in relationship with the purchaser as an agent acting on behalf of users.

The question when evaluating the reforms is to what extent the conditions are present in order to function as a quasi-market for employment and training services. Table 4 summarises the criteria which can be derived from the identified problems for each of the five conditions.

The conditions and related criteria or indicators for the functioning of quasi-markets for social policy can be explained as follows:

- (a) market structure; various providers and purchasers must be present; new providers must gain easy access; as well as existing

Table 4: **Conditions and criteria for the functioning of quasi-market for social policy**

Conditions		Criteria
Market structure	Competition	Many providers and many purchasers
	Exit	Easy entry for new providers
Market prices	Possibility of exit	Prices are formed directly
	Equal treatment to providers with different ownership status	
Information	Accurate	Access to cheap and accurate information
	Complete	Formulation of contracts
Transaction costs	Certainty	Monitoring of quality
	Transparency	Moderate transaction costs for purchaser and provider
Motivation	Motivated by financial considerations (providers)	Tendering and contract conditions
	Motivated by needs of users (purchasers)	New entrants
Cream skimming	Needs of users determine services/ treatment	Balance profit/not for profit motivation
		Purchaser acting on behalf of users
Measures to avoid creaming		
Source: Based on Le Grand and Bartlett (1993); van der Veen (2000); Trommel et al., 2001		

monopoly structures, space must at least be created for a counterbalance, for example by also allowing local authorities to function as purchasers;

- (b) information; easy access to information about costs and quality must be available. This can be organised by setting up one umbrella service to act as an information manager for all the required aspects;
- (c) transaction costs; transactions are complex and multi-dimensional and require a certain specialisation; together with a high level of uncertainty (e.g. concerning what will be needed in the future) this can lead to high costs. Transaction costs should not be higher than the costs of the previous system;
- (d) motivation; providers must also be guided by financial interests and motivated to anticipate market signals. Purchasers must be motivated to work towards the welfare of the user (above their own organisational interests);
- (e) cream skimming (creaming); the user or client receives a free service, but discrimination can occur with respect to 'more expensive' users. This can be remedied by linking the need to the remuneration offered, for example by giving poorly educated people greater weighting in standardisation or financing.

5.3. Available evaluation studies

For empirical analysis of quasi-market reforms in this paper, we can rely upon own research carried out for the Ministry of Employment in the Flemish Community of Belgium and the OECD (Struyven and Steurs, 2002; Struyven et al., 2002). The research consists of literature review and qualitative case studies, based on an analysis of written source material and 47 interviews with important players in the four countries ⁽²⁾. These sources provided the main input for the formative evalua-

tion in this paper. In addition to this research, the analysis has been extended with a review of more recent documents and evaluation studies (up to December 2002) and specific information gathered by telephone interviews or e-mail correspondence. Studies have been screened on available findings for both formative and summative evaluation, though much more information about the detailed functioning and impact evaluation of the quasi-market seems to be available from Australia. As the only parallel to Australia is the emergence of a private reintegration market in the Netherlands, the overview in this paper is mainly based on empirical evidence from both countries. However, one must consider that there is, so far, no comparable monitoring and evaluation available from the Netherlands ⁽³⁾. For these reasons the summative evaluation review is based on evidence from Australia only. Of particular interest in this paper is the importance of training and the impact of quasi-market reforms in job brokerage and reintegration services on the provision of training. The evidence on this point from the Australian Job Network draws heavily on a memo prepared for us by the Australian Evaluation and Programme Performance Branch of the Department ⁽⁴⁾.

5.3.1. Australia

At the introduction of Job Network there was considerable public interest in the process for reviewing such a radical shift in delivery arrangements. A comprehensive evaluation strategy was developed which involved three reporting stages, recognising that the reforms would take several years to develop and mature, as well as an independent review ⁽⁵⁾.

The first two stages of the official evaluation reported in May 2000 and May 2001. These reports were largely based on performance under Job Network's first contract. Stage 1 focused on the implementation of Job Network and its early operation, while stage 2 examined equity of

⁽²⁾ Face-to-face interviews conducted in the period between June and October 2001 with the authorities, the market players, the social partners and independent experts.

⁽³⁾ After publication of this paper, perhaps progress will have been made with regard to monitoring and evaluation data.

⁽⁴⁾ With thanks to Mr Philip Gatenby of DEWR.

⁽⁵⁾ There has also been a Probity Adviser's report on the first Job Network tender in 1999, as well as a performance audit undertaken by the Australian National Audit Office, which we do not consider further here. Finally, the OECD review on Australia's labour market policies was for the most part devoted to the Job Network (OECD, 2001).

access to assistance and outcomes, early indicators of the impact of assistance (on the likelihood of leaving income support) and regional performance (DEWRSB, 2000a; 2000b). The third report, published in May 2002, provides information on the lessons learned from evaluating Job Network over the period of its operation since May 1998 (DEWR, 2002b). In assessing effectiveness, the Stage 3 evaluation examined the sustainability of employment outcomes, the impact of the major services in improving employment prospects, how well Job Network is endorsed by its clients, its responsiveness to special needs and its macro-economic impacts. The evaluation also considered the relative contribution of factors that can explain the level of effectiveness achieved, and how Job Network has progressed in the light of its design principles. This paper is mainly based on the evaluation findings of the third evaluation report.

The evaluation strategy also included an independent review of the policy framework underpinning Job Network. This *Independent Review of the Job Network* was conducted by the Productivity Commission and was published in June 2002 (Productivity Commission, 2002). This commission had to examine critically and comment on the framework for delivering labour-market assistance, including the application of the purchaser-provider model to employment assistance, the roles of the relevant players, areas where the model could be improved and the possible scope for applying the model to other types of services. In conducting its review, the Commission had regard for research already undertaken, but also drew on community input and called for submissions from interest groups and the broader community.

Apart from the report published by the Productivity Commission, most evaluation studies have been carried out within government by the Department of Employment and Workplace Relations (DEWR) as the department responsible for administering the Job Network. This gives rise to concerns about the independence of the findings. To date no external scientific evaluations have been carried out. A general problem here is the lack of data. The Department is not willing to release data for scientific research, mainly citing reasons of confidentiality. This explains the lack of hard empirical evidence so far on the effective-

ness of Job Network. The Productivity Commission also notes in its report that 'Increased transparency would also help to increase the exposure of the Job Network to the discipline of external scrutiny by independent researchers'.

5.3.2. The Netherlands

In the Dutch reintegration market, the Minister has undertaken to monitor developments closely through an annual evaluation of SUWI (Eerste Kamer, 2001). Evaluation plays a prominent part in the Dutch process. As is true for policy in general, the Dutch government devotes considerable resources to various types of internal and external evaluation research.

In contrast with the evaluation concerns expressed by the Dutch authorities in general, to the end of 2002 only monitoring studies are available on the emerging reintegration market, with evaluation studies of tendering procedures. External studies on the emergent market structure contain a process evaluation of the market arrangements in practice. These studies were commissioned by the Department of social Affairs and Employment, first published in 2000 (Dijk et al., 2001) and updated in 2002 (Mevisen et al., 2002). In addition, central indicators are being developed by the benefit institution (UWV), the municipalities and the advisory body (RWI), but there is no common integrated statistical system (see below). UWV tendering has been evaluated, in particular to consider the general efficiency of tendering and contracting (Cordia et al., 2001; LISV, 2001; UWV, 2002). Furthermore, an economic overview of the evolution of the Dutch reintegration market since the beginning of the 1990s has been published (Groot et al., 2002). This study, carried out by the University of Amsterdam, was commissioned by the Department of Economic Affairs and mainly based on available documents and interviews. Finally, a country specific study of the Netherlands was published in the special Job Network issue of the Australian Journal of Labour Economics (Struyven and Steurs, 2003). This study was based on original field research (document analysis and interviews) concerning the implementation in 2000-02 (Struyven and Steurs, 2002).

5.4. Summary

In this chapter we have derived two sets of criteria from the theory of quasi-markets expounded in Le Grand and Bartlett. The first set of criteria relates to the functioning of the quasi-market; the second is derived from the objectives of market forces policy. The first set constitutes formative evaluation, the latter summative evaluation. In the next chapters, the

evidence concerning the quasi-market reforms in the employment and training sector are structured following this evaluation framework. The main questions for formative evaluation concern the rationale in design and implementation. To what extent are the quasi-market arrangements consistent with theoretical principles? To what extent are they relevant to the needs of the labour-market actors? For summative evaluation, the main question concerns the impact on quality and effects: do the reforms meet the objectives?

6. Formative evaluation

6.1. Market structure

Probably the most important structural conditions for quasi-markets to function properly relate to market structure. In the cases we consider in this paper, market competition has been introduced by means of open tendering. This requires a separation between purchaser and provider, not only as an institutional setting, but also in the minds of the actors involved. Subsequently, in order to gain the advantages of competition, various providers (supply side) and purchasers (demand side) must be present. Tendering procedures and contract conditions may have an important impact on the degree of provider competition. Because of the importance of a level playing field between providers, we also pay attention to the position of the former Public Employment Service. A final issue in relation to the market structure which we address is the importance of price competition as opposed to quality competition.

6.1.1. Separation between purchaser and provider

Tendering requires a strict separation between the purchaser and the provider, known as the purchaser/provider split, in order to avoid judge/judged situations.

In all cases studies, the split between purchaser and provider is inherent in the model. Nevertheless, the new institutional setting does not mean that the culture of tendering is automatic. This is most clearly noticeable in a situation where market forces are a recent phenomenon. Experiences in the Netherlands show that, in addition to structural changes, a change in mentality is also needed with respect to dealing with preferential relationships from the past. This is perhaps one of the most stubborn transitional problems which a government has to face when shaping market forces.

6.1.2. Provider competition

Provider competition is necessary if the advantages of tendering are to be grasped (Fay, 1997;

Derksen et al., 2000). A competitive supply side will, however, often be missing when market forces are introduced into a field where services have traditionally been organised chiefly by the government, as in the case of employment placement and reintegration. The design of the tendering procedure and the content of the contracts can help in the development of a sufficiently broad providers' field (Domberger, 1998). After a discussion of these factors, we take a closer look at price competition in relation to quality concerns. Table 5 presents the tendering procedure in the four countries in diagram form.

6.1.2.1. *The relationship between tendering procedures and provider competition*

The first factor which influences the development of provider competition is the scope and duration of the contracts. On the one hand, contracts can be too small to stimulate the supply side and, on the other hand, the scope of the contract can be so large that only large providers can participate. For this reason, activities can be grouped together or split up into smaller units to interest enough providers, or collaboration can be permitted. This is the case in the Netherlands ('tranches' in the UWV tendering) as well as the UK (different New Deal tenders). The length of the contract is also important. A longer duration enables the providers to recoup their investments. For instance, in Australia some potential new participants found the term of 19 months under Job Network 1 too short, which led them to decide not to participate. However, too long a term can prevent potential new participants from entering the market. For this reason, the County Labour Market Board in Sweden, for example, rarely makes use of the possibility open to it of extending contracts.

Another factor which influences the number of providers is the specificity of investments which are required to provide the service. For instance, the provision of activation and reintegration

services requires the deployment of highly specific human capital. Organising the provision of certain training courses is also associated with considerable start-up costs. Experience in Sweden shows that competition exists particularly in training courses which can be offered by the providers in return for relatively low fixed costs (administration, IT, etc.), while in capital-intensive sub-markets (welder, mechanical engineer, etc.), far fewer entered the market.

A third factor is related to quality requirements. A great deal of attention is – rightly – devoted to the quality aspects of reintegration services and training. The quality requirements

which are imposed can, however, become de facto obstacles to entry to the market. For example, the experience requirement in the Dutch UWV tendering process means that new participants do not stand a chance (see below).

6.1.2.2. Price versus quality competition

If the quality of a product or service to be provided can be unambiguously determined, the price can be used as the only selection criterion in the tendering process. This condition is under no circumstances fulfilled when it comes to employment placement and reintegration services. As well as the price, the proposed

Table 5: **Tendering procedure**

	Australia	The Netherlands	United Kingdom	Sweden
Type of tendering	Tendering with one central purchasing organisation	UWV tendering: tendering with central purchasing organisation	New Deals: tenders with local purchasing organisations Employment Zones: tendering with central purchasing organisation	Tendering with local purchasing organisations
Process	Open tendering in one round	UWV tendering: qualified open tendering in two rounds Municipal tendering: still being developed, so far few open tenders	New Deal Programme: qualified open tendering in two rounds	(Usually) open tendering in one round
Price versus quality competition	Job Network 1: if quality criteria are met, primarily price was important, fixed prices for Intensive Assistance Job Network 2: 75 % quality, 25 % price, with minimum prices for Intensive Assistance	UWV tendering: price and quality competition, but price never most important criterion	Fixed prices, therefore 100 % quality competition	Price and quality competition, with price weighted at 30-40 %
Transaction costs	UWV procedure: Considerable; leads to review of procedure with proposal for 'roll-over' of contracts	New Deals: limited by two-phase procedure but registration and offer per individual contract	Duration of limited by introduction of two-phase procedure	three months regarded as too long

quality must also be a major selection criterion. Relevant and proven experience and expertise are one possible criterion. But this may imply another trade-off with the intensity of competition, because too much emphasis on proven expertise rules out the entry of new players to the market. Moreover, past performance may be unreliable, because outcomes achieved may depend on many, and sometimes random, factors. It may be very hard to differentiate objectively between providers that have different capacities to generate outcomes. If bidders cannot adequately signal and pre-commit to deliver higher quality, then the major basis for discriminating between tenders bids will be price. Accordingly, while price competition at the bidding stage can be an important feature of purchaser-provider arrangements, it cannot be given full reign when the purchaser does not know how much they are buying or the overall quality of the service.

In each of the countries examined, it is clear that the introduction of market forces cannot be related to the introduction of price competition, as quality plays an important role. In the UK, fixed prices are used in tendering, for each of the three programmes under consideration. Since the selection criteria are related to the quality on offer, we can talk of 100 % quality competition. In Australia, the price criterion was given a weighting of 25 % in the second and most recent

tendering process, while the quality offered was given a weighting of 75 %. In principle, this enabled higher prices for Job Network providers that might achieve greater outcomes than the average. Free prices apply for job matching and job search training while for intensive assistance prices were subject to competitive bidding, but a floor price was set. The floor price protected quality, but to a large degree it also set the price for many Job Network providers as they bid down to the floor. Therefore, the Productivity Commission, in its independent review of the Job Network, recommends that the department set default prices for Job Network services (Productivity Commission, 2002). This has actually also been proposed by the government in its discussion paper on the new arrangements for the third tender to come (DEWR, 2002a). In the Netherlands, free price competition exists, but the price is not the most important criterion in any single contract. When tendering for training in Sweden, the price is given a weighting of 30-40 % and the price is often used as the first selection criterion.

6.1.3. Player's market

In this section, we deal with the actual developments which have taken place on the supply side in the countries under consideration, including a discussion of the position of the former public provider. Table 6 gives an overview of the (changes in the) situation in the four countries.

Table 6: **Current situation**

	Australia	The Netherlands	United Kingdom	Sweden
Number providers	– 205 organisations at 2 114 sites – NGOs, private sector, public sector	– UWV contracts: 33 main contractors – Total market >650 companies – Private sector, municipal companies	– New Deals: 2 500 contracts – Total market: 1 800 approved organisations	Very extensive market
Public provider: (former)	– Public company, fully conforms to market	– In transition as public company with guaranteed budget until end 2001, then privatised	Public Employment Service as benefits agency and provider of first route	– Lernia (formerly AmuGruppen) = public company, privatised in two phases (1986, 1993), fully conforms to market
Public Employment Service	– Market share sharply reduced – Long tradition of contracting out to private players	– Gradual reduction of compulsory purchase from Public Employment Service	phase (variable)	– Declining market share

	Australia	The Netherlands	United Kingdom	Sweden
Other public providers	None	Municipal companies, educational establishments	Educational establishments	(municipal) educational establishments
NGOs	– Sharply increasing market share – Previously already active as provider – Professionalisation, scale increases and concentration – Joint ventures	NGOs are classed as public organisations	Less known about them, do not really stand out	– Wide variety of players – Trade union organisations – Former employees of AmuGruppen – Entry and exit depending on scope of government budget
Private providers	– Sharply rising market share – Previously already active as provider – Few new members	– From pure organiser to provider – Increase in scale and alliances, e.g. with insurers as well as niche players	Employment Zones: conscious choice for close involvement of private players (Working Links = public-private joint venture)	
General trends	– Market concentration – Relatively few entrants	– Purely private market with sub-markets (medically-oriented, work-oriented) – Driven by insurers – Very mobile market	– Developing market – Concentration movement through tendering of complete packages	Entry in urbanised regions and for non-capital-intensive training
Market share by public provider	Employment National: 8 %	KLIQ: 17 %, UWV market, 50 % municipal market (expected)	Employment Service as principal	Lernia (formerly AmuGruppen): 25 %

6.1.3.1. Concentration and increases in scale

What developments actually took place in the countries considered? In Australia, 306 providers were selected in the first round of tendering (Job Network 1), who offered their services at over 1 000 sites. The second tender reduced the number of candidates to 205, while the number of sites rose to 2 114. Providers had thus become larger and the supply side more concentrated. In Job Network 2, the four largest providers together accounted for roughly a third of the market for each of the services provided. Job search training is the least concentrated market segment, but concentration increased most significantly for these services (see Table 7). The concentration in intensive assistance fell sharply

because Employment National withdrew almost completely as a provider in this market, while it held 42 % of the market under Job Network 1.

In the Netherlands, the number of contracted providers increased from 33 after the first UWV-tendering to 41 after the second UWV round. This meant there was a reduction in the concentration, though this remained high. The four largest providers had a combined market share of almost 59 % after the first round, and this was still more than 47 % after the second round (calculations based on data provided by LISV). Note that the UWV-market represents only a sub-market within a broader market structure that is constituted by the UWV, the municipalities and the employers (Struyven and Steurs, 2003).

Table 7: **C4 concentration ratio***

(%)

	Job Network 1	Job Network 2
Job Matching	32	35
Job Search Training	21	31
Intensive Assistance	65	34
* The C4 concentration ratio corresponds with the sum of the market shares of the four largest providers. The market share is calculated on the basis of the contracted volume.		
Source: Davidson, 1998; DEWRSB, 2000a.		

In order to guarantee a varied supply, rules are sometimes applied regarding a minimum number of providers. For instance, the department responsible for allocating contracts in Australia assumes that at least 5 providers are present in each region and that none of the providers will be allocated more than 50 % of the total contracted volume in each region. In the Netherlands, a rule states that if a contract is split up into several 'tranches', a reintegration company cannot be awarded more than one tranche. In the UK too, when contracts are allocated, a sufficiently broad mix of market players is ensured.

6.1.3.2. *Difficult entry to the market*

Overall, it also seems that entry to the market has remained relatively limited, especially by the large providers. Under Job Network 1 in Australia, one in five providers was new, while under Job Network 2 fewer than one in six of the providers was new. Most providers, even the genuinely private parties, had already been involved as providers previously – after tendering – in the labour-market programmes that can be regarded as the forerunners of Job Network. In the Netherlands, 18 new companies entered the market in the second tendering round, but together they only obtained a market share of about 10 %.

This limited entry to the market is related, among others, to the selection criteria, which

include the expertise and experience of the potential providers. The Australian National Audit Office observed in its analysis of the results of the first tender that organisations which at that time were in receipt of funding for providing case management or other, related employment placement services, achieved a much better result than other candidate providers. This is clearly illustrated by the figures in Table 8.

While the average success rate after the first round of tendering was 30 %, it was significantly higher for already established providers (43 %) and much lower for new providers (14 %). In spite of this, almost 45 % of the candidates were new providers.

In the Netherlands, requirement for experience excludes new entrants to the market. It is envisaged that 20% of the contracts can be allocated without tendering, for example to new players who put forward an innovative approach. In Sweden, new players are entering the market, particularly in the urbanised regions and for training courses which are not capital-intensive. This can be explained by the widely fluctuating scope of demand, which means that new players do not run the risk of investing in the setting up of expensive training programmes.

Table 8: **Success rates of established versus new providers (Job Network 1)**

	Established providers	New providers	Total
Number of applicants	569	447	1 016
Successful applicants	243	63	306
Success rate	43 %	14 %	30 %
Source: ANAO (1998)			

6.1.3.3. *The position of the former public employment service*

The introduction of market forces has had a significant influence on the status and role of the (former) public employment service. In Australia, the Netherlands and Sweden, the introduction of market forces was (partly) responsible for reforms to the public employment service because it is important to create a level playing field in which providers are treated equally.

In Australia, the Commonwealth Employment Service was closed down and a new company set up: Employment National. This new public company still only has the function of provider of employment placement and reintegration services, where it has to compete with other providers. The remaining tasks of the Commonwealth Employment Service were assumed by the department and by Centrelink. Whereas Employment National still played an important role in Job Network 1, its market share has fallen sharply in Job Network 2; for intensive assistance it has actually fallen from 42 % to 1 %. The transition to a public company which has to compete on an equal footing with other providers has been very difficult, involving substantial staff cuts. As a result of the sharply declining market share in the publicly financed placement and reintegration market, the company recently developed activities for the private market. But the Minister for Finance and Administration has announced that the marketable parts of Employment National will be sold and the remainder of the company wound up on 30 June 2003.

In the Netherlands and Sweden the public provider was not abolished, but reformed. In the Netherlands, KLIQ is being privatised and all the reintegration activities of the employment service are being made its responsibility. KLIQ is also a purely commercial enterprise, which no longer has any privileged relationship with the government. Privatisation has been accompanied by staff cutbacks and financial problems. A similar story can be told about AmuGruppen, the government provider of training courses in Sweden. This organisation was privatised in two stages, into a company operating fully in line with the market and whose shares are now held 100 % by the Swedish government. AmuGruppen – now known as Lernia – also saw its market share in the publicly financed training market decline sharply,

leading to reductions in staff numbers. An increasingly large percentage of its turnover now comes from activities for the private sector.

6.1.4. **The position of the purchaser**

6.1.4.1. *Degree of monopoly at the purchaser side*

A crucial determinant of the operation of market mechanisms in social service delivery is the degree of monopoly that characterises the client side. In a quasi-market it is usually not the direct user who exercises the choices concerning purchasing decisions; they are delegated to a third party such as a government department or authority in employment and training (Le Grand and Bartlett, 1993). In some cases, there is a multitude of purchasers and a multitude of providers. The opposite situation is a tendency towards bilateral monopoly, with a single purchaser and a single provider. In between lie situations in which there are few purchasers and many providers, or few providers and many purchasers (Walch, 1995).

Practice in the various countries varies, the position of the purchaser being important. Australia and the Netherlands are the two extremes: Australia has only one purchaser; in the Netherlands, the tendering practice tends towards a quasi-market with a large number of buyers and providers. In between lies the UK system, which is characterised by many providers and few purchasers (Employment Service and Department of Work and Pensions) which operate several tender processes. Table 9 represents the different market types.

The Dutch situation is related to the importance attached to the municipalities for labour-market policy, as well as to employers, who have been made financially responsible for sickness and disability benefits (Struyven and Steurs, 2002; 2003). With regard to the population of job seekers, the purchase structure is left up to the autonomous government institution (UWV) and to the municipalities. The differentiated purchaser structure has a greater chance of leading to a quasi-market with a multiplicity of purchasers and suppliers. However, Struyven and Steurs point out that in 2001–02 this market structure had not yet introduced much differentiation on the purchaser side. The most important purchaser, UWV, still very much has the char-

acter of a monolithic regional body inviting tenders; purchasing by municipalities is still at an early stage (Struyven and Steurs, 2002). In general the municipalities have not yet built up enough expertise for their own purchasing policy. Municipalities are used to operating as 'benefit factories' and now they have to transform themselves into 'reintegration directors'. Often, there is a lack of knowledge of both the client and the labour market. One key difficulty is that it is not yet fully in the interests of the municipalities to have an open market. They are playing a double role because they have public companies themselves providing jobs under the Job Seekers Reintegration Act and the Social Employment Act. In principle, these municipal companies have to compete in a way that is fully in accordance with the market, which means that they are also registered for VAT. It seems that a purchasing policy is only a possibility for the larger municipalities. Only a minority of municipalities organise an open bidding process, while others operate more selective tendering procedures (Mevissen et al., 2002).

6.1.4.2. Degree of centralisation at the purchaser side

Another determinant in the functioning of market mechanisms is the degree of centralisation. To avoid problems of coordination while developing market processes, governments tend towards the establishment of tendering frameworks, uniform procedures and contracts, performance measurement systems, etc. This tendency reflects pressure for centralisation, whereas the specification and fine-tuning of tendering conditions and contracts tend towards decentralisation. According to Walch, there is a threefold distinc-

tion in the organisational structure of a market model: the corporate core responsible for strategy and policymaking; the client side for services responsible for setting and monitoring standards; and the service provider, who actually delivers services (Walch, 1995).

The early experiences in the four case studies indicate that there is no single way of resolving the tension between centralisation and decentralisation. The tendering mechanism can be structured from the central policy level or a smaller territorial policy level. In that respect, the four case studies examined exhibit major differences. In Australia, competences are strongly concentrated, with the central department in Canberra acting as sole purchaser as well as policy-maker. This contrasts with the decentralised model in the Netherlands and Sweden, where the regional level (County Labour Boards in Sweden) or the local level (municipalities in the Netherlands) have an inalienable role as purchaser and principal. The situation in the Netherlands (UWV tendering) and the UK (New Deals) is closer to the centralist model, although regional departments have more impact on the tendering process.

Decentralised purchasing organisations offer a dual advantage: they are closer to the field so that they are better able to assess needs and, since the market has more buyers, the relationship between buyers and providers is more balanced. The use of a uniform tendering procedure and model agreement is important: this contributes not only to greater transparency of the tendering process, but also to lower transaction costs. Each of the three countries mentioned attempts to standardise the procedures.

Table 9: **Market types based on the number of providers and purchasers**

		Providers	
		Few	Many
Purchasers	Few	Bilateral monopoly/oligopoly	Monopsony/oligopsony e.g. the Australian Job Network
	Many	Monopoly/oligopoly	Free (quasi-)market e.g. the Dutch reintegration market

Source: Based on Walch, 1995

6.2. Information and transparency

Following the quasi-market theory, markets only operate efficiently when both sides of the market have access to cheap and accurate information, particularly concerning the costs and the quality of the service concerned. Le Grand and Bartlett (1993) point out that 'Providers must be able to cost their activities so as to be able to price them appropriately. Purchasers must be able to monitor the quality of the service they are purchasing, so as to limit the opportunity for providers to reduce costs by lowering quality'. The information deficiencies stem from the outcome-based focus of the market model. An outcome-based approach does not need to specify what processes are used by providers to obtain jobs for their clients, as competition over time should shift providers to the best methods (Productivity Commission, 2002). However, it may still be in the interests of the purchaser to discover what processes are actually most effective and to diffuse them among their suppliers. Furthermore, information deficiencies may underlie problems in price competition and job seekers' choice. Finally, transparency is considered to be an important policy principle, where the provision of public funds is involved (Productivity Commission, 2002).

6.2.1. Lack of transparency

In the case studies of Australia and the Netherlands, there is poor information about what works in practice. This is seen as a disadvantage of payment based on outcomes (Productivity Commission, 2002). Publicly available information about what providers actually do with their clients is relatively poor. This is especially so with the content and intensity of services, such as the importance of training where the available monitoring and evaluation documents offer few details. Information sharing and coordination between providers are limited. Job seekers struggle to make choices because of a general lack of transparency in the system.

In the Netherlands, little attention is yet being paid to addressing the problem through data collection (Struyven and Steurs, 2002). In general, computerisation is lagging behind. The provision of information and facilities to monitor reintegration efforts have not been able to keep up with the rapidity of the changes. On the purchaser side, the

UWV tendering procedure has created more transparency in their corresponding market segment (Cordia et al., 2001). A uniform tendering procedure, model agreement and predetermined selection and award criteria also increased the transparency of the procedure compared with the first tender and contributed to lower transaction costs (UWV, 2002). One crucial question is whether the UWV tendering procedures are focused in a one-sided way on the new inflow of unemployment and disability insurance benefit claimants. Placement and drop-out figures have an important part to play in the payment of fees; it might, therefore, be expected that there would be much more transparency in the validity and reliability of the placement figures used. In the municipalities market segment, several monitoring initiatives have been set up in the context of different programmes and funding mechanisms, but without any coordination. According to an external monitoring report ordered by the government, the municipalities are suffering from an overdose of monitoring and evaluation (Mevissen et al., 2002). This can be explained by the absence of an overall integrated employment and client information system.

In contrast to the Netherlands, Australia already has an employment information and service delivery system. This system provides extensive information about Job Network members (including information on their performance) and services to job seekers. Despite this, a general lack of transparency is mentioned by the Productivity Commission's report and recognised by the Government's response (Productivity Commission, 2002; DEWR, 2002b). The Commission recommends that provision of information to job seekers about Job Network and the associated referral system be enhanced to allow job seekers greater scope for informed choice.

6.2.2. Contractual compliance monitoring in Job Network

The Australian Department acting as a purchaser has in place an extensive regime for monitoring and quality assurance. It makes use of codes of conduct, declarations of intent and intensive assistance support plans as well as the Job Network contracts themselves. According to DEWR, the Job Network Code of Conduct is the central feature of consumer protection under Job Network. The aim of the Code is to produce the best outcomes for job seekers and employers by

developing a high-quality, continuously improving service that engenders ethical behaviour between all parties. All Job Network members are required to meet the minimum standards set out in the Code, which forms part of their contract with the Commonwealth. Organisations tendering for intensive assistance in the second round were required to submit declarations of intent that described the range of services they would provide to job seekers. The declaration of intent forms a part of a provider's contract with DEWR ⁽⁶⁾.

Many providers perceive increasing compliance burdens that direct them away from their main goal of placing disadvantaged job seekers in employment (Struyven et al., 2002). The use by providers of the Integrated Employment Service is very complicated. Clearly, there will always be a need for a level of compliance monitoring that can ensure that outcomes paid for are in fact delivered, and detect and deter unintended behaviours by providers. But too strict a monitoring regime will impose a high administrative burden on both the department and providers, and has the potential to reduce flexibility and deter innovation. The *Productivity Commission Report* raises the idea of an independent agency to oversee the Job Network and department's management of it. Such an agency could address issues of transparency and accountability, and ameliorate any problems of power imbalance and distrust between the Job Network players and the department. For the Commission this is a conditional recommendation 'if significant problems do continue' under the new contract period (Productivity Commission, 2002). The Commission also recommends that all data on Job Network programmes that is not confidential be made available for independent scrutiny by other researchers.

6.2.3. Performance measurement in Job Network

To assess the performance of providers, the Australian Department has designed a star rating

model, to be used to renew contracts and to pay outcomes. The model uses a set of performance indicators and associated weightings based on the performance indicators outlined in the Job Network contracts ⁽⁷⁾. A provider's actual performance is assessed against its expected performance, adjusted to take account of variations in client mix and local labour-market conditions. Scores are distributed between one and five stars such that 70 % of providers in a region are rated at three stars or better. The star ratings system has occupied a central place in the operation of the market.

The star rating appears to have increased Job Network's focus on securing outcomes for job seekers. Stakeholders have both welcomed and criticised this approach. Supporting a focus on outcomes is based on two arguments:

- (a) a good way to secure a lasting job is to accept any job in the short-term. In other words, there are more likely to be lasting benefits for job seekers if the provider focuses on securing a job placement rather than on spending time building up skills, or progressively addressing any barriers to employment. But critics claim that short-term employment may not necessarily in the long term interest of some disadvantaged job seekers (DEWR, 2002c, pp. 152-153);
- (b) a greater weight to outcomes for more disadvantaged job seekers combats incentives for providers to work most intensively with the easiest to place candidates.

Nevertheless, many participants are critical of the model, mainly because of concerns that it lacks transparency and measures performance on too narrow a basis or with too much error (Struyven et al., 2002; Productivity Commission, 2002). Job Network providers do not know the specification of the model. They are also mixed in their understanding of where the 'goalposts' are. There are cases where the rules have been changed after providers have geared their businesses to meet the original rules. This in turn has

⁽⁶⁾ Each provider must draw on its declaration of intent to draft an intensive assistance support plan that specifies the activities and services it will provide to each individual job seeker remaining on their caseload 13 weeks after they commence. While the intensive assistance support plan is not legally binding between the provider and the job seeker, DEWR checks the provider's delivery of activities as specified in the support plan as a part of its contract monitoring activities.

⁽⁷⁾ For example, in the case of intensive assistance, the current weightings are as follows: 40 % for interim outcomes (13-week outcomes) as a proportion of commencements, 30 % for final outcomes (26-week outcomes) as a proportion of commencements, 10 % for the proportion of participants placed in a job and 20 % for the proportion of interim outcomes going to disadvantaged job seekers.

impacts on business confidence, certainty and investment. The Commission's report recommends that full details of the star rating model be made publicly available, including periodic assessments of its technical validity. 'Transparency will help alleviate some of the distrust of the model and allow better models to be constructed in the future'. The evaluation also recommends that all declarations of intent (and similar documents that specify services to be provided to job seekers by individual providers) as well as information on contracts (nature, extent and cost; compliance monitoring) should be made public (Productivity Commission, 2002).

Another critical point is what is actually measured by the star rating. Placement outcomes cannot be assumed to be a measure of the overall quality of service provision. Performance is only one aspect of quality. The key problem is how to define and measure an outcome.

6.3. Transaction costs

The transaction costs associated with setting up a tendering process, both for purchaser and provider, are often underestimated (Domberger, 1998). Transaction costs for purchasers include, for example, the costs of writing and publishing the specifications, evaluating the offers, drawing up model contracts, completing negotiations and monitoring the agreements. Typically, all this requires a great deal of senior management time and a level of expertise that may not always be present within the body which will be responsible for the tendering process. Providers also incur transaction costs, for instance when preparing their bids. In each of the case studies, the transaction costs emerged as an important area of attention.

In Australia, the preparation of the two previous tenders, evaluating the offers and contract processing required the deployment of huge numbers of people and vast resources by the purchaser. But most providers in the Job Network also regard competitive tendering as expensive and time-consuming. The tender documents are very complex, as are the contracts. The tender process — preparation of tender bids, assessment of bids and finalisation of contracts — disrupts service flows noticeably

during the second contract. Renewing the market through a global tender also contributes to business uncertainty among providers and confusion among job seekers and employers. The government has therefore proposed that around 60 % of the existing capacity be rolled over under a third contract for the best-performing providers, with the remaining share being subject to tender bidding (DEWR, 2002a). The contract rollover will partly avoid the high disruption costs to job seekers of Job Network 2, but since many providers may still be bidding for capacity at the Employment Services Area level, the contracting transaction costs may not be appreciably lower than previously. This suggests that the costs involved in a hybrid rollover/tender model mean that it too is not sustainable.

An alternative, proposed by the Productivity Commission (though not followed by the government), is to use licensing as the basis for entry into the industry (Productivity Commission, 2002, p. XLI). Licensing would be dependent on a provider meeting certain minimum standards and agreeing to various contract provisions (such as a code of conduct, undertaking assessment of job seekers and whatever activity testing was agreed). The standards set would be similar to those established by the department for pre-tender qualification. Renewal of the licence would be tied to performance, so that the vital role performed by existing arrangements in driving poorer suppliers from the market would be maintained.

In the Netherlands and the UK, attempts are made to limit transaction costs by working with a procedure in two rounds. In a first stage companies come forward as candidates for the stated target groups. In order to be invited to take part in the second round, predetermined, objective selection criteria have to be met. In a second stage, the purchaser invites the selected reintegration companies to submit a tender. Contracts are awarded on the basis of predetermined award criteria. This implies savings for both the principal and the potential providers. Another advantage of this selection procedure is that a larger number of providers will be interested in applying, given the relatively limited efforts which have to be made to submit an offer. However, if only a limited number of providers can actually submit an offer after selection, this could result in higher prices because the competition is limited.

6.4. Motivation: the combination of input and output financing

The balance of motivation is affected by the introduction of market incentives (Le Grand, 2000). In order to influence the provider's behaviour and to tackle information asymmetries, much attention has been paid to the payment structure. The relationship between the principal and the providers is characterised by a number of information imbalances which influence the payment structure used (Canoy and Vollaard, 1999). These information asymmetries relate in the first instance to the factors which determine the costs of reintegration: the technology and the efforts that have to be made by the employment placement consultant (provider). The technological capacity of an employment placement consultant relates, for example, to its access to information networks, its knowledge of the demand for labour or its knowledge of reintegration opportunities. Clearly, these providers will know more about their technological capacity than the government. The efforts of the employment placement consultants also determine the extent to which they use their capacities efficiently, for example the amount time devoted to each client. The government could monitor these efforts. But the result of the efforts in the form of the number of reintegrated benefit claimants will be more easily seen by the government, making an abstraction of the measurement problems and of whether the output is as desired. Information asymmetries may also arise because the employment placement provider has a clearer insight into the profile of the unemployment pool for which a reintegration programme is being developed. That provider can then focus on the most easily placed candidates (creaming).

6.4.1. Input and output financing

These information asymmetries create problems when it comes to paying providers. The government will want to structure funding arrangements in such a way that providers are encouraged to achieve efficient and effective reintegration. The theory puts forward a combination of input and output financing as a solution: the employment placement consultant receives an 'initial fee' in

advance per person to be placed and a 'placement fee' afterwards if the efforts made result in placement of the job seeker in a job. Through the output financing part, the providers are encouraged to provide services which produce results. The input financing results in a fair spread of risk between the principal and the contractor.

It is clear from the case studies that the payment structures used in practice often fit in with the theoretical solution. The providers receive an initial fee as soon as a number of minimum requirements are satisfied (for example, the formulation of an 'activity agreement' for intensive assistance in Australia). In the event that the counselling results in an 'outcome', typically defined as placement of the job seeker in a job, the providers receive a placement bonus. Since the placement of a job seeker in a permanent job is better than placement in a temporary job, the placement bonus is sometimes divided up, with a share only being paid when the job seeker has been working for a certain number of weeks. For instance, an 'interim placement fee' is awarded for intensive assistance in Australia if the job seeker remains in work for 13 weeks and a 'final placement fee' after 26 weeks. A comparable system applies to the employment zones in the UK. More details of the payment structure are summarised in Table 10.

6.4.2. The Dutch payment system

The Dutch payment system deserves special attention for two reasons. The first is that providers must 'guarantee' a placement percentage (= result commitment) in their offers, which also applies as a selection criterion. Second, providers may earn a bonus if they do better than they promised. The payment consists of three parts: a fixed payment, a placement fee and a bonus, and depends on the following parameters:

- N = the number of clients commenced
- C = the contracted price per client (specified by providers in their bids)
- f = the fixed payment percentage (specified by UWV in its call for tenders) ⁽⁸⁾.
- P_g = the 'guaranteed' placement percentage (specified by providers in their bids)
- P_r = the realised placement percentage

⁽⁸⁾ The more difficult the target group, the higher the fixed payment, in order to combat creaming.

Table 10: **Payment structure and competition**

	Australia	The Netherlands	United Kingdom	Sweden
Input versus output financing	Job matching: only output financing Job search training: combination of input and output financing Intensive assistance: combination of input, interim outcome and final outcome financing	UWV tendering: combination of input and output financing, plus bonus if performances are better than contractually stipulated	Combination of input and output financing	Only input financing
Measures to combat creaming off	Providers cannot refuse job-seekers Intensive assistance: job-seekers divided into various categories	Large number of target groups, individually tendered Minimum placement percentages per target group The more difficult the target group, the higher the fixed compensation	New Deals: providers cannot refuse job-seekers; large number of packages with different prices Employment zones: output financing dependent on length of unemployment	
Choice for job-seekers	Theoretically yes, in practice barely	For larger target groups, choice between several providers	Some job-seekers do have options	Job-seekers are referred on by Public Employment Service Office

The ‘fixed payment’ (like the commencement fee in Australia) is paid for every client who actually starts a programme, independent of the placement percentage. It is equal to $N \cdot C \cdot f$. The ‘placement fee’ is outcome based. It is equal to $N \cdot C \cdot (1-f) \cdot P_r$. The final part is ‘the bonus’, which is also outcome-based but which also depends on the comparison between the realised placement percentage and the guaranteed placement percentage. It is equal to $N \cdot C \cdot (1-f) \cdot P_r \cdot (1-P_g)/P_g$.

Providers are rewarded if they do better than they are contractually obliged to. This is not the case within Job Network, where there is no bonus payment. The bonus formula also implies that a

higher guaranteed placement percentage has a negative impact on the size of the bonus. But the invitation to tender specifies minimum placement percentages which vary for each target group. Higher minimum placement percentages apply to the target groups which are easier to place in work, in order to counteract creaming.

The total payment is equal to $N \cdot C \cdot (f + (1-f) \cdot P_r / P_g)$. This formula clearly shows that if the realised placement percentage is higher than the guaranteed placement percentage (i.e. $P_r > P_g$), a company receives more than the price submitted ($N \cdot C$) ⁽⁹⁾; but the reverse is true if a provider fails to place as many clients as contracted (i.e. when $P_r < P_g$) ⁽¹⁰⁾.

⁽⁹⁾ Suppose a company offers to help 100 clients to find a job ($N = 100$) for a price of EUR 1 000 per client ($C = \text{EUR } 1\,000$). The guaranteed placement percentage is 55 % ($P_g = 0.55$). The fixed payment for the target group is 40 % ($f = 0.4$). If the company is able to place 60 clients ($P_r = 0.6$), it will receive a total payment of EUR 105 455.

⁽¹⁰⁾ If the company can only place 50 clients in a job ($P_r = 0.5$), the total payment will be EUR 94 545.

6.4.3. Output financing and training

One important condition for the use of output financing is that the intended result can be clearly and objectively defined. This is not easy for training. The purchaser might only pay the training provider if the job seeker successfully completes the course or, to direct incentives towards training which really helps job seekers to find a job, if the completed training results in employment.

There are, however, a number of disadvantages to such an approach. Training providers would be encouraged to keep job seekers in their course longer than may be necessary such as in cases where the job seeker was offered a job before the end of the training. Providers would also be financially penalised by early leavers.

If a job seeker finds a job after completing a training course, it is not clear how it can be proved that finding the job was the consequence of the training. This helps to explain why Sweden only envisages input financing for training providers. Nevertheless, the expert group on efficient training programmes for the unemployed is of the opinion that the National Labour Market Board should try to develop mechanisms where the payment for providing vocational training is linked with performance standards (Näringsdepartementet, 2000). The two possibilities they suggest are to link the payment to individual knowledge goals or to link a bonus to the percentage of participants who have employment after the training programme.

Another disadvantage of output financing of training services is that providers may not be willing to take many risks and may invest too little in expensive training for the job seekers they are counselling, given that they are not certain of being able to recover these costs through a placement fee.

The absence of specific training funding in the outcome-based payment structure seems to be a problem in Australia, where there is widespread – if anecdotal – evidence that the amount of training provided has been massively reduced by the Job Network providers (see below). However, training can be individually compensated, as in the Netherlands.

6.4.4. Impact of the outcome-based focus on provider behaviour

The star rating model, as applied in the Australian Job Network, has a significant impact on provider behaviour. A profit maximising contracted agency will aim for the best outcomes as measured by the contract indicators, but not necessarily the best overall outcome. In any system designed around outcome payments, providers will work to the incentive structure built into the contract. Ideally the contract should incorporate all important outcomes. This is not always easy to achieve: ‘[...] competition for “stars” [is] continuing to take the focus away from “what is best for the client” and instead emphasising “how can I get a payable outcome?”’ (Mission Australia in Productivity Commission, 2002, p. 3.13). This goes against the integrity and the spirit of the Job Network (Ibid., 11.20). There are many possible manifestations of this in the Job Network. Where there is any discretion in taking or subsequently excluding clients, incentive payments based on simple performance outcomes may lead to the taking on of those job seekers most likely to achieve payable outcomes (creaming). In addition, providers may avoid job placements that last less than 13 weeks for intensive assistance clients or less than 15 hours for job matching clients, these being the relevant thresholds for outcome payments specified in the contract. Job Network providers may also have incentives to manipulate the system by ‘parking’ (providing little assistance to) clients with low employment probabilities, creating temporary artificial jobs that maximise placement payments, and rotating people through them. Job outcomes *per se* may underemphasise or ignore the wage level, duration or quality of a job.

In the Netherlands, a debate is going on whether to increase the weight of the placement fee and to move on to a ‘no cure no pay’ system, at least for the job seekers who are relatively easy to reintegrate ⁽¹¹⁾. This may cause several disadvantages: providers will raise their prices, they will meet problems of liquidity, and they will invest less in the hard-to-place job seekers or in excessive training for the job seekers. Until now, training has been individually compensated for in the Netherlands. In the future this may change since the idea

(11) An estimated 70 % of the total number of job seekers.

is to finance training only if the provider manages to find a job for the job seeker (SZW, 2002).

6.5. Creaming and parking

Competition within quasi-markets in combination with output financing can give rise to 'creaming': the providers select the job seekers who are the easiest to place, in order to increase their profits. Job seekers at the greatest distance from the labour market are left out in the cold. While creaming refers to the practice of selecting job seekers who are the easiest to place, parking refers to the practice of providing little assistance to those with either insurmountable or high barriers to obtaining work, or who have low likelihoods of achieving payable outcomes. But low levels of assistance are not necessarily attributable to the provider alone: parking might also be a behavioural response by demotivated job seekers (self-parking).

The contracts and payment structures as used in the countries under consideration contain a number of elements which are explicitly intended to counteract these phenomena. First, providers typically cannot reject the job seekers who are referred on to them: this is the case in Australia but not in the Netherlands.

Furthermore, it is also usual to classify the job seekers into various categories, depending on their distance from the labour market, and to adjust both the amount and the relative weight of the commencement fee and the outcome fee. For instance, job seekers for intensive assistance in Australia are divided into two categories, with the (minimum) price being higher for the category of those most difficult to place in work. However, in Job Network 2, the number of categories was reduced from three to two, which increases the chance of creaming within the categories. In the Netherlands, 22 target groups were covered by the UWV tendering process in 2001, a number of these being grouped by sector. Tendering took place at the level of the (partly) sector-specific target groups. This is intended to provide customers with more tailor-made services, as well as counteracting creaming within excessively large groups of people. In the (target group-specific) New Deal programmes in the UK, the services put out to tender are divided into a large

number of packages depending on their duration and content, which restricts the possibilities for creaming within those packages.

In addition, contracts with providers in the Netherlands include a type of result commitment in the form of a placement percentage (see above). The minimum placement percentages vary for each target group: higher minimum placement percentages apply to the target groups which are easier to place in work.

Finally, the initial fee for the target groups which are more difficult to place is higher than the placement fee (see above). A form of positive creaming applies within the Employment Zones: the placement fees are higher for job seekers who have been out of work for over three years, so that providers can concentrate their efforts on those most in need.

The problem of parking could be overcome by moving to a fee-for-service system in which any of a wide range of service types would simply be paid for by government when delivered to job seekers. In that case, parked job seekers would receive services because it would then be in the interests of providers to maximise service delivery, rather than outcomes. However, over-servicing and poorly targeted service delivery are well-known risks of a fee-for-service system.

By defining the target groups as specifically as possible and also by region, the aim is to deliver more tailored services to job seekers. This can also be seen as a way of preventing creaming within wider populations. Finally, it fosters smaller specialised companies to enter the market. Nevertheless, a majority of providers in the Netherlands consider the large number of target groups as negative. According to the main Dutch purchaser, less differentiation in target groups may be expected during the third round because most of the job seekers have recent work experience and can be easily reintegrated (UWV, 2002).

From a formative evaluation point of view, it can be stated that the case countries are fully aware of the danger of creaming, and have therefore incorporated a large number of stipulations and refinements in the design of the market system. In the summative evaluation the question is whether these adjustments do actually avoid creaming, parking and related problems of inequity.

7. Summative evaluation

7.1. Efficiency

There is widespread evidence from a number of authoritative studies that market systems are the most efficient (Savas, 2000). Savings seem most likely to be made in the case of simple and repetitive services such as refuse collection and cleaning. In more complex services, such as health and social care, there seem to be few savings to be made (Deakin and Walch, 1996). However, some scepticism is justified, as ‘many studies rely on assertion or on surveys of management perception’ (Ibid.). For employment and training services there is mixed evidence about possible efficiency gains.

For the Netherlands, there is as yet no evidence of increased efficiency in the reintegration market compared with the preceding period. Cost savings do not play a role, as the overall budget for reintegration activities was increased during the reform period (Mevissen et al., 2002). The total volume of business is estimated to be EUR 770 million, accounting for 160 000 job seekers (Groot et al., 2002). Prices are rising, with an average of EUR 4 700 a trajectory in 2002, compared with EUR 3 500 in 1998 provided by the former public employment service *Arbei-*

dsvoorziening (Ibid.). So one cannot yet conclude that the market model in the Netherlands has become more cost-efficient, since prices are higher compared with the pre-market period.

Evidence from the Australian Job Network seems to confirm the expected gain in cost-efficiency. According to the final evaluation report, the introduction of Job Network has resulted in a substantial reduction in the cost of achieving employment outcomes (DEWR, 2002c). Job Network costs per employment outcome ⁽¹²⁾ have been the lowest achieved in the past decade: at about AUD 5 000-6 000 since mid-1998, compared to between AUD 10 000 and AUD 16 000 in the mid-1990s and AUD 8 000-9 000 in the early 1990s (Table 11).

In 2000–01, costs per employment outcome were AUD 560 for job matching, AUD 1 390 for job search training and AUD 5 440 for intensive assistance. The costs for the second and third were dramatically below those of the comparable programmes that they replaced: respectively, AUD 2 600 for job clubs and AUD 12 100 for the group of programmes replaced by intensive assistance. The substantial efficiency gains made by Job Network have also translated into improved cost-effectiveness, measured in terms

Table 11: **Cost-effectiveness of Job network**

Period	Cost per employment outcome (AUD)	
Early 1990s	8 000 – 9 000	
Mid 1990s	10 000 – 16 000	
Late 1990s	5 000 – 6 000	
Type of service	Working Nation	Job Network
Job matching	–	AUD 560
Job search training	2 600 ^(a)	1 390
Intensive assistance	12 100 ^(b)	5 440
^(a) Job Club		
^(b) Average for seven programmes replaced by Intensive Assistance		

⁽¹²⁾ These are gross outcomes.

of costs per net impact (DEWR, 2002c, p. 126-127) ⁽¹³⁾. The improved cost-effectiveness of Job Network is largely a function of greater efficiency rather than greater effectiveness, since the net impact of Job Network services was found to be similar to the net impact of the previous labour-market programmes (see below).

The independent review by the Productivity Commission gives some additional figures which allow an evaluation of the efficiency of Job Network (Productivity Commission, 2002, p. 5.24). For the four years from 1996/97 to 1999/00, the Government allocated more than AUD 5.4 billion to fund all labour-market assistance and entry-level training programmes; AUD 1.6 billion for 1996/97, and an average of AUD 1.3 billion per year thereafter (of which Job Network programmes have averaged about AUD 0.75 billion per year for the past three years). This compared to annual average funding in 1994/95 and 1995/96 of AUD 3.7 billion.

7.2. Responsiveness and choice

A key indicator of Job Network's responsiveness is its capacity to target services to job seekers. Therefore, we first examine application of the job seeker classification instrument which streams job seekers to the different services provided within Job Network. A subsequent issue of responsiveness we pay attention to is the choice of provider made by job seekers. Finally we address two other indicators, i.e. job seeker and employer perceptions of quality, by a short presentation of the results of the customer satisfaction surveys.

7.2.1. Application of the job seeker classification instrument

If job seekers can be reliably identified and streamed to the service that is most likely to make a difference to their employment prospects, resources can be targeted to where impact is highest. Under current arrangements, effective targeting relies to a large extent on the application and accuracy of the job seeker classification

instrument. This aims to identify job seekers who are most likely to remain unemployed or become long-term unemployed.

The application of the instrument has proved troublesome (Productivity Commission, 2002, p. 9.3-9.13). A survey of providers in 2001 found that 21 % of job search training providers and 4 % of intensive assistance providers thought that over half the job seekers referred to them were inappropriately classified (DEWR, 2002c, p. 106). In qualitative research some providers indicated that many of the job seekers referred to them should have been referred to either a more intensive service or to services outside Job Network. One of the problems mentioned is that job seekers do not disclose personal information because they do not understand the purpose of the questions in the instrument. A review of the job seeker classification instrument interview process also identified some measurement errors and inconsistencies in its application. Several factors were found to either involve a degree of ambiguity or make it difficult for job seekers to answer accurately.

Overall, we can conclude that the job seeker classification instrument is welcomed with much criticism. In the implementation it is perceived as an instrument for rationing to detect those job seekers who have the 'capacity to benefit', while previous employment and training programmes were accessible for everyone (Struyven et al., 2002). For a quasi-market, it is very important to have a confident classification instrument, as the level of fees to providers is dependent on the classification of the job seeker. If the job seeker is classified too low, the provider will have less incentive to put in effort (parking).

In the Netherlands, an analogous instrument is used, called the 'chance-meter'. Job seekers who are capable of finding a job on their own with a minimum of support are allocated to stream 1, while those who are not at all ready for a paid job in the open labour market are allocated to stream 4. Job seekers allocated to streams 2 and 3 are expected to need a 'trajectory' (a period of individual case management, perhaps seen as a pathway through distinct phases such

⁽¹³⁾ According to the Productivity Commission, the cost reductions in terms of cost per net impact could also be the result of the imprecision with which the small impacts of labour market programmes are measured (Productivity Commission, 2002, p. 5.24).

as training, work experience and job search) lasting less than one year, or one to two years, respectively. The chance-meter does not say anything about the content of the trajectory that a person needs. It is not yet clear how far this instrument succeeds in predicting the cases in the right way.

7.2.2. Choice of provider

In normal markets, choices by consumers have a major effect on company performance and incentives for provision of quality services. For this incentive effect to work, consumer choices must be informed and must systematically favour better over worse enterprises. However, the auto-referral system, quantity caps and poor information provision to job seekers mean that most job seekers are assigned randomly to Job Network providers. On the basis of administrative data, Centrelink estimated that only about 20 % of job seekers chose their own provider, with the remainder being allocated a provider through the automated referral system. In addition, where choice is exercised, it does not appear to be highly informed. From DEWR's job seeker satisfaction survey, it appears that of those making a choice, convenient location seems to be the most important criterion for judgement. This raises the issue of whether and how to increase the scope for choice in a way that rewards better performers with a greater flow of referrals. Many observers question (rhetorically) whether job seekers, even with information, are able to make a real choice (Struyven et al., 2002).

In the Netherlands, job seekers have been given more options in the second tender. For target groups of a specific minimum size, contracts are signed with multiple reintegration companies. In many cases a job seeker can then choose from a number of different companies, together with his or her contact person from the UWV. But it remains unclear how this will work in practice due to the lack of informed choice.

7.2.3. Client satisfaction

According to the client satisfaction surveys of the Australian Department, the majority of job seekers appear satisfied with the services they receive from Job Network. Among job seekers surveyed in 2001, 74 % reported that they were either satisfied or very satisfied with the services. Satisfaction levels were over 80 % for job search

training and intensive assistance. Among the reasons job seekers reported for being either satisfied or very satisfied with Job Network services were that the assistance helped them stay connected with the workforce and they were 'treated like an individual'. For job matching, however, the level of satisfaction with services was only 56 %. The Productivity Commission is quite sceptical about these figures: 'It is common for satisfaction surveys to reveal satisfaction levels of around 80 %, irrespective of the topic' (Productivity Commission, 2002, p. J.1). Similarly, the Productivity Commission found out that in the 2001 survey, satisfaction is inversely related to the star rating of providers.

Similar results are found in the perceptions of employers. According to the department, employers who use Job Network rate the service highly. Among employers surveyed in 2001 who had used Job Network in the previous 12 months, 90 % of them were satisfied or very satisfied (increasing from 84 % in 1999) (DEWR, 2002c, p. 113). But the Productivity Commission, commenting on the 1999 survey results, notes that the corresponding satisfaction levels reported for other recruitment agencies was 90 %, despite the fact that only 15 % of employers reported having to pay for Job Network services compared with 62 % for other agencies (Productivity Commission, 2002, p. 6.24). The most often cited benefits from using the Job Network related to screening of applicants and saving time interviewing and finding applicants. The most often cited poor aspects were poorly trained/incompetent staff and poor service. Nevertheless, the Australian Chamber of Commerce and Industry reported that its 1999 survey of 1 200 members found 'that employers were overwhelmingly satisfied with the performance of the providers, particularly in comparison with the former Commonwealth Employment Service' (Productivity Commission, 2002, p. 6.23).

In the Dutch case, for the time being no satisfaction data are available.

7.3. Quality

Using market forces is demonstrably more efficient than past models, then, but is it possible that this is achieved through a lower service level

or poorer quality? In what follows, we discuss the kind of services offered to job seekers within Job Network, with special attention paid to the amount and type of training provided.

7.3.1. Services of the Australian Job Network

Providers within Job Network have the flexibility to tailor the services to the needs of the job seekers referred to them for assistance. The rationale for this is that decisions about how to help the unemployed to overcome their barriers to employment are best made at the local level between a job seeker and their employment consultant. There is, for instance, no prescription within intensive assistance for the delivery of certain types of labour-market assistance.

Consistent with this philosophy, the government does not regularly collect data on the types of services delivered within intensive assistance. This information has to be obtained from surveys. The 2001 Job Network participants' survey, for example, indicated that most job seekers (over 90 %) within intensive assistance received job search assistance in the form of help with CVs and interview preparation. Just over half the job seekers reported receiving some personal assistance (counselling and presentation advice). Over 40 % of job seekers received financial assistance, including assistance with fares and the cost of special equipment. Only 8 % of job seekers obtained work experience (DEWR, 2002c, p. 88).

In the Netherlands, at the end of 2002 the first quality labels, developed by the branch organisation Borea, were awarded to 27 reintegration companies. It is anticipated that 13 performance indicators may in the near future generate data about throughput, success ratio and drop out ratio, as well as satisfaction levels with clients and employers.

7.3.2. Training within Job Network

Of special interest for this paper is the impact Job Network had on the amount and type of training provided to job seekers.

7.3.2.1. The amount of training

The 2001 Job Network participants' survey distinguished between training in job search skills, such

as help with CVs and preparing for interviews, and more substantial job skills training, such as IT training, language/literacy/numeracy training and training linked to a specific job (job-specific training). Training of this type was reported to have been received by 23 % of intensive assistance participants surveyed in 2001. Of all the episodes of training reported in the 2001 survey, 41 % lasted less than a week, while for 32 % of episodes the duration was between two and eight weeks.

How does this compare with the period before Job Network? There has been a shift towards 'job search' skills and a reduction in the amount of short-term training. During the first half of the 1990s there were over 100 000 annual commencements in labour-market programmes which offered short-term training. Since the introduction of Job Network indications are ⁽¹⁴⁾ that around 64 000 job seekers participate in short-term job skills training annually as part of intensive assistance. But over the same period there has been an increase in participation in job search assistance provided through Job Network's job search training service. In the early 1990s there were fewer than 30 000 annual commencements in this type of assistance. The number of commencements in job search assistance was over 85 000 in 1997 and over 80 000 in 2001.

The significant role job search assistance plays in intensive assistance is further evidence of a shift towards job search activities. The evaluation of Job Network found that over 90 % of intensive assistance participants received assistance with job search from their providers (counselling and help with preparing CVs and for interviews).

The move away from short-term training is consistent with a 'work first' approach and is supported by evidence on the relative effectiveness of different types of labour-market assistance. One of the motives for reforming the labour-market programmes and introducing Job Network was the observation that training was failing as a labour-market instrument. The government criticised the lack of results of existing labour-market programmes: '[...] few seemed to make any significant or lasting difference in getting unemployed people back into regular employment' (DEWRSB, 2000a). Training

⁽¹⁴⁾ On the basis of the number of job seekers starting intensive assistance in 2001, the reported level of training is equivalent to around 64 000 participants receiving training in job skills.

programmes were criticised in particular, such as the SkillShare programme, which offered short-term training opportunities. Services offered such as job search assistance, job referral and continuing personal support contributed more significantly to the outcomes of the programmes than services that provided structured training, either in combination with these and other forms of assistance or on their own (DEET, 1993).

7.3.2.2. *Outcome based payment and incentives to provide training*

The fact that job seekers are offered training to only a limited extent, ⁽¹⁵⁾ is attributed mainly to the higher-risk nature of these investments because providers cannot be certain that an investment in training will result in a job for the job seeker. In addition, the bias in the fee structure towards employment outcomes may, in some cases, favour the pursuit of short-term employment outcomes over training and education which may, in the longer-term, be of greater benefit to job seekers.

It can be concluded that, before the introduction of Job Network, where job seekers were too easily placed into training programmes without this improving their chances of finding work, at present too little is being invested by providers within Job Network in training for job seekers. The introduction of training accounts as part of the Australians working together welfare reform in July 2002 can be seen as a tacit concession in this regard. These schemes enable eligible mature (aged 50 and over) and indigenous job seekers to undertake work-related training to the value of AUD 800. Job seekers can choose from courses on the Australian Training website. In order to participate, training must be accredited and delivered by registered training organisations. Other reforms in the third Job Network contract in July 2003 also led to an increase in payments for providing services to job seekers, without these payments being made dependent on a job outcome. For example, a job seeker account will be introduced which will be available to eligible job seekers who will typically enter intensive support services after three months in unemploy-

ment. A Job Network member may spend this money (on behalf of the job seeker) on training, including that supplemented by a training account.

7.4. Net impact and equity of the Australian Job Network

Employment outcomes are the ultimate touchstone for the summative evaluation of the introduction of market mechanisms in employment and training services. A number of questions can be identified here. The first is whether the concentration on outcomes does not lead to a very strong focus on short-term results in the form of helping job seekers back into employment, possibly at the expense of results in the longer term (sustainability of outcomes). The second question is whether the new arrangements succeed in achieving better employment outcomes than before (before-and-after comparison). The third question ensues from the creation of a market with different providers: to what extent do the differences between providers lead to differences in employment outcomes (comparison between providers)? Finally, we can ask to what extent the outcomes are also achieved for the more difficult categories of job seekers, and then specifically in relation to the incentive structure for providers which focuses on outcomes (equity and impact of the fee structure on equity).

7.4.1. Gross outcomes

It is generally acknowledged that gross outcome measures mean very little in terms of evaluating the impact of employment and training programmes, because many of the outcomes would have occurred anyway (Productivity Commission, 2002). Therefore, impact evaluations should focus on net impacts, in terms of how many people moved off benefits to go into paid work as a result of the assistance and training provided. As stated by the Productivity Commission, this is an exercise beset by data-related and methodological problems,

⁽¹⁵⁾ It is important to note that a substitution effect may have occurred from federal level to state level, whereby although less training is offered in within the Federal Job Network programme, it may be that more training programmes have been initiated by the state governments.

requiring estimation of a variety of programme effects, including:

- (a) dead-weight losses: programme funds spent on job seekers who would have found work anyway;
- (b) compliance effects: where some job seekers referred to a programme find work or otherwise leave benefits before the programme starts (especially when participation is compulsory as in Australia);
- (c) attachment effects: where programme participants reduce their job search activity for the duration of a programme;
- (d) displacement effects: where a programme participant displaces someone out of the programme.

It is not the purpose of this paper to deal with these problems in depth. Our focus is on the impact of the functioning of the market for employment and training services. Since the case of Australia is the best documented in terms of evaluation studies, our discussion is restricted to this one case.

7.4.2. Sustainability of outcomes

As regards the gross outcomes of Job Network, placement figures are available, measured three

months after the end of the relevant measure. The department collects these data via a written survey, the post-programme monitoring survey, which has been held since the mid-1980s. The advantage is that these data are collected continuously and provide an insight into short-term results. According to the department, the lower response rates among people in work lead to an underestimate of the positive outcomes. The findings are presented in Table 12.

The table shows that the percentage of participants in job search training who are in employment three months after leaving the scheme rose over the period considered from 36.2 % to 43.2 %. Half of these are in full-time employment, the other half are in part-time jobs. The percentage of participants in intensive assistance in work three months later was 38.6 % in the period 2000-01. This percentage is significantly higher than in the preceding year, though lower than in 1998-99. Roughly 40 % of those in work have a full-time job; the remaining 60 % are in part-time positions. If outflow to training programmes is also regarded as a positive outcome, the positive outflow percentages (which can not be completely equated to employment commencement percent-

Table 12: **Gross outcomes of Job Network**

	1998-99	1999-00	2000-01
Job search training			
Employed	36.2	38.2	43.2
Full-time	18.3	19.9	21.5
Part-time	17.9	18.3	21.6
Further education	13.6	12.6	12.9
Positive outcomes	45.5	46.7	52.1
Intensive assistance			
Employed	39.6	32.7	38.6
Full-time	18.8	14.4	16.3
Part-time	20.8	18.3	22.3
Total	100.0	100.0	100.0
Further education	7.7	8.0	7.7
Positive outcomes	45.8	39.3	44.6
NB: Outcomes related to job seekers who ceased assistance between April 1998 and March 1999, April 1999 and March 2000, and April 2000 and March 2001, measured three months after leaving assistance. Positive outcomes include employment and education/training outcomes, but are not the sum of these two outcomes because some job seekers can achieve both an employment and education outcome.			
Source: Based on DEWR (2002c)			

ages) are slightly higher, at 52.1 % for job search training and 44.6 % for intensive assistance (figures relate to 2000-01).

In the view of the department these outcomes are quite sustainable, which is one of Job Network's main objectives (DEWR, 2002c). Across the three main services, about four out of five of those in jobs three months after assistance were employed in the longer term:

- (a) for job matching, 83 % of job seekers who were employed three months after placement were employed 12 months later;
- (b) for job search training and intensive assistance, 82 % and 76 %, respectively of those employed in the short term were also employed after eight months.

The evaluation results of the department also indicate that, for many job seekers, the quality of outcome (in terms of both income and skill levels) improves over time, supporting the government's view that for some job seekers a low-quality initial job can provide a stepping stone to a better quality job. The sustainability of outcomes also appears to hold generally for most job seekers, including those who are more disadvantaged. There is, however, a group who return to income support over time. Among job seekers referred to intensive assistance who left income support, for example, about 30 % had subsequently returned to the income support register 15 months after their referral.

7.4.3. Before-and-after comparisons

Direct comparisons between programmes over time are constrained by the fact that programmes may have different objectives and target different job seeker groups, and may operate under different labour-market conditions. An internal evaluation study is available in which a before-and-after comparison is made between Job Network and six labour-market programmes that operated in the mid-1990s (DEWRSB, 2001). Following the methodology of matched comparison groups⁽¹⁶⁾, the study found a net impact of around 10 percentage points for intensive assistance and 3 percentage points for job search training. This compares with the results of

previous programmes in the mid-1990s (though individual programmes varied widely).

7.4.4. Provider comparisons

A central feature of a quasi-market is the existence of different providers (instead of a monopoly). The department's evaluation reports some findings based on a comparison between high-performing providers and low-performing providers in terms of job seekers' perceptions of the quality of the staff, and the types and quality of assistance provided (DEWR, 2002c). The evaluation found that the best providers seem to be more proactive and more responsive to the needs of both job seekers and employers. Furthermore, the more successful providers are the ones who are more likely to offer training, covering both job search and job-specific skills.

7.4.5. Net impact and equity

The evaluation by the department presents estimates of the net impact of job search training and intensive assistance on the employment prospects of participants. Unlike previous net impact studies, the estimates in the third evaluation report use a new methodology with matched comparison groups that attempts to measure the cumulative effect of three elements: compliance, programme and attachment effects. The net impact in this study was found to be modest: the Job Network programmes generate a very small net positive effect for participants (job seekers who commenced employment): 8.3 percentage points for job search training and 0.6 percentage points for intensive assistance. The low result for intensive assistance can be explained by the attachment effect of the length of time job seekers spend on the programme. Finally, the evaluation attempts to measure the compliance net effect, which was estimated to be almost two percentage points.

According to the department, the findings need to be interpreted very carefully because of the inherent difficulty of forming a control group where there is universal entitlement to assistance (DEWR, 2002c, p. 3). The net impact measures do not reflect a pure comparison between inter-

⁽¹⁶⁾ Comparison group of job seekers who had not been referred to or participated in the programmes in the previous six months, matched on age, gender and duration of benefits. Both programme and comparison group sample were large (26 000 job seekers each for the intensive assistance programme and 5 000 job seekers for the job search training programme).

vention and no intervention; rather, they compare an intervention with a combination of no intervention and other forms of assistance. In addition, the estimates do not measure the improved employment prospects of participants in assistance who have not yet secured a job. Finally, positive net impacts are a static measure of effectiveness. Even if modest, these benefits compound for successive groups of participants, producing more significant effects on the level of disadvantaged job seekers when considered over a longer timeframe. Nevertheless, the publication of the results gave rise to public concern, including in Parliament, about the effectiveness of Job Network. The department's impact evaluation also prompts scepticism about the appropriateness of the matched control group methodology (Productivity Commission, 2002). No attempt is made to control for unobservable personal factors in influencing success in the labour market, such as motivation, demeanour and attitude to work. This problem is known as selection bias.

Despite the tentative nature of the findings, some evidence is available concerning the equity of outcomes. The Department's evaluation has found that more disadvantaged groups benefit most from the programme, even though these job seekers have lower gross outcomes from assistance than other job seekers. This is the case for the net impact of the following groups: mature job seekers (aged 45 and over), the low-educated (less than 10 years in education), and the long-term unemployed (12 months and over). This corresponds with international evidence of net impact measurement of labour-market interventions. However, it is found that financial incentives focusing on employment outcomes have an adverse effect on the equal treatment of all job seekers. Based on a modelling of the incentive effects of the current fee structure, the evaluation concludes that the financial incentives in the fee structure make it rational for providers to focus assistance on those who require only a limited level of assistance in order to obtain outcomes.

8. Conclusions

8.1. Assessment of evaluation strategy

Discussion of evaluation of quasi-market reforms has shown that evaluation findings about process and impact are limited. This results from the inherent difficulties in defining and measuring outputs, net impacts (by using control groups) and quality issues. Furthermore, an evaluation of reforms should include before-and-after comparisons, which suppose appropriate longitudinal research to distinguish rigorously between the effect of market forces and other changes that may have taken place simultaneously, such as changes in the benefit system or in the labour market.

None of the countries under consideration have a systematic evaluation strategy for the new arrangements, except in Australia, but research to date has been monopolised by the central government. Outside researchers cannot get access to original data, nor they can get an insight into its processing. Implementation of the reform by providers, job seekers and employers remains a black box due to the confidentiality of the data. This leads to the conclusion that official evaluation studies are somewhat problematic and

function mainly as a justification of policy decisions. In this respect, the independent review by the Productivity Commission was a breakthrough in governmental research strategy. This review is not only based on data and findings by the department, but also on the experiences of all stakeholders, and has ended in a critical debate. Evaluation of quasi-market arrangements should include all stakeholders in a more interactive way and keep at least part of the evaluation work for external independent researchers.

8.2. Main findings

The experiences from the Australian and other quasi-market reforms support the view that the market model in providing reintegration services is superior to that operating previously. However, there is no hard empirical evidence for this. The main findings of the summative evaluation are summarised in Table 13.

The Australian case provides a very mixed picture. The most significant gain is that the market system leads to more cost-effective performance: services cost less, with a more or less similar net impact as previous programmes and

Table 13: **Empirical evidence concerning the impact of quasi-markets in Australia (1998-2002) and the Netherlands (2000-02)**

Criteria	Trend	Main findings Australia	Main findings the Netherlands
Efficiency	+	Gain in cost-efficiency Gain in cost-effectiveness (<greater efficiency)	Not yet efficiency gains
Responsiveness	+/-	Targeting with job seekers Classification instrument troublesome Local needs	Screening and streaming with 'chance-meter': predictive value? Tends to involve regional needs
Quality	+/-	Erosion of the level of services, especially job specific training Scepticism about satisfaction levels	No empirical evidence available Quality labels to 26 companies (end 2002)
Choice	-	No real choice for clients (<20 %) No informed choice	No informed choice No empirical evidence available
Equity	-	No strong empirical evidence Low net impact (0.6 %; 8.3 %) More disadvantaged benefit most	No empirical evidence available

adequate responsiveness and quality. Weaker points are the limited choice for job seekers and erosion of the level of services, especially with regard to job specific training. However, these findings about net impact and quality are tentative and should be added to other findings concerning the amount of training services, the financial incentives and their role for disadvantaged job seekers. It was found that the financial incentives focusing on employment outcomes have an adverse effect on the equal treatment of job seekers. The incentive structure favours the pursuit of short-term employment outcomes over training and educa-

tion which may, in the longer-term, be of greater benefit to job seekers.

In the Netherlands, it is not yet clear how well the arrangements to foster responsiveness and choice between providers may work. The Dutch reintegration market tends to be self-regulating on quality and, at this moment, there is no evidence of increased efficiency nor equity.

Despite these mixed findings, more positive evidence is available with regard to the formative evaluation, i.e. the evaluation of the conditions that are crucial for the functioning of quasi-markets (Table 14).

Table 14: **Empirical evidence concerning the conditions for quasi-markets in Australia (1998-02) and the Netherlands (2000-02)**

Condition	Trend	Main findings Australia	Main findings the Netherlands
Market structure	+	Many providers, but one central purchaser Difficult industry entry Administered prices and floor prices Clear distinction between purchaser and provider Deficiencies of the Job Network market: concentration, large companies favoured, instability	Many providers, increasing number of purchasers (sub-markets) Low barriers to industry entry, however experience as a selection criterion Free prices Different tendering culture between purchasers Deficiencies of the UWV-market: concentration, large companies favoured, instability, market of organisers
Information	+/-	More transparency of procedure, not yet transparency of the market Extensive regime for monitoring and quality assurance, increasing compliance burdens to providers Integrated labour-market information system, but data conceived as confidential Star rating model lacks transparency	More transparency of procedure, not yet transparency of the market Poor monitoring information, problem of data collection Development of a quality label by the sector itself
Transaction costs	+/-	Tendering causes high transaction costs Roll over of 60 % contracts under the third round	Annual tendering causes high transaction costs Lowered by a uniform two-stage procedure and model agreements
Motivation	+	Combination of input and output payment Selection criteria based on price (25 %) and quality (75 %) Training not separately financed Star rating model focussed on 'payable outcomes', leading to risk-averse behaviour Tendency to give more choice to job seeker	'No cure less pay' Training separately financed Tendency towards 'no cure no pay' leading to risk-averse behaviour Tendency to give more choice to job seeker
Cream skimming	+/-	Providers cannot reject job seekers, some adverse effects Job seekers divided in different (broad) categories; higher price for the most difficult category	Number of elements to counteract, some adverse effects Number of categories; tendency to less target groups

Typical problems concerning information and transparency as well as transaction costs can be overcome by refinements in structure, tendering mechanisms, implementation and monitoring arrangements. In this respect, the experiences are not similar, as shown in Table 14. The differences between Australia and the Netherlands represent contrasting market types, with both advantages and disadvantages, as summarised in Table 15.

One advantage of the Australian model is that direction from one centre leads to a high level of uniformity. This promotes the transparency of the market and clarity of legislation. But single principalship places the Australian department in an omnipotent position. As monopsonist, it has a monopoly on the tendering policy. This is further reinforced by the high level of policy centralisation.

The Dutch model is characterised by more equal relationships between buyers and providers. More room is available for the various players, as well as for the more specialised market segments. A pluriform principalship therefore promotes the position of the customer, who has more chance of finding a service provider he likes in a more varied and differentiated players' market.

However, the evidence discussed in this paper shows that both models also have disadvantages. Australian principalship operates in a hierarchic/bureaucratic way as a result of its large number of rules and control. In the interviews,

this was characterised as a legalised market or 'lego'-market (Struyven et al., 2002; Productivity Commission, 2002). In addition, the Job Network system is increasingly driven by performance measurement using star rating, which is disputed in the field. The star rating itself is not rejected but, as a result of the lack of open communication about the criteria used and its method of calculation, it meets with much resistance. Both factors influence each other. The distance to the field is further reinforced by the need to keep everything controllable. At the time of the research, the market was cloaked in an atmosphere of mistrust concerning the Job Network. The question arises whether this encourages the players to take risk-avoiding measures, which is precisely what the government wants to counteract by the way it acts. A second question is whether this might prejudice the final results.

The Dutch market type suffers from a lack of transparency. A difference in tendering culture still prevails between the principals. Experiences yet have to show how the transition to a business-like principalship can succeed in a model where the purchasing function still remains fairly close to the implementation domain. The principals can also be faced with undesirable side-effects in the behaviour of the providers, such as shopping and cross-subsidising between different public purchasers.

Table 15: **Advantages and disadvantages of two market types based on tendering practice in Australia and the Netherlands**

	Advantages	Disadvantages
: Monopsony Australian type	Directed from one centre Steering towards outcomes Transparency of the market Clear rules Legally sound	Obsession with performance measurements Hierarchic/bureaucratic Dominated by rules and control ('legalised market') Large distance from field Atmosphere of mistrust; lack of open communication
Quasi-market: Dutch type	More equal relationships between buyers and providers Promotes varied playing field More specialisation possible More communication Strengthens position of customer (options)	Less transparent market Differences in tendering 'culture' Shopping for providers Cross-subsidies possible between different public assignments

The differences between the two types can be brought into focus in yet another way. Both tendering practices are founded on a mixture of hierarchic and market relationships (Thompson et al., 1991; Parsons, 1995; Bouckaert, 1998). Typical of the hierarchic relationship is that the dominant player directs the other actor on the basis of rules and standards. The task imposed on the directed actor can change at any moment. In a market relationship, the mutual relationship is managed by the market mechanism between demand and supply. A market relationship is characterised by more balance between equal players. In Australia, a hierarchic relationship is expressed in many ways: the price-regulating role of the government, the fact that the providers of intensive assistance and job search training also have to assume job matching services, etc. The Australian system has more in common with a holding in which various companies operate and where the central management controls through a mix of hierarchy and market. In a less centralised system with more buyers, more room is created for allowing the mix of hierarchy and market to interfere through a cooperation relationship, based on agreements.

8.3. Lessons for other countries

While comparing reform in the two leading countries, Australia and the Netherlands, one should

take into account the institutional and labour-market contexts. It is clear that market forces are not the only factor affecting the output of employment brokerage and reintegration activities. However, the operational characteristics of the system do matter. Despite the different context of the Australian welfare state compared with the continental or Scandinavian model, on the practical level of matching supply and demand and helping different categories of job seekers, all countries would appear to be confronted with the same problems. Quasi-market experiences can help in seeking new solutions. These elements can deal with structural conditions, such as the purchaser-provider split, the importance of a decentralised approach, organisational formality, the incentive structure, and implementation conditions such as job seeker classification and referral instruments, an integrated labour-market data system or quality management. Even without implementing a quasi-market model, all these conditions play a role in the search for better practice in employment services. A central role is played by the incentive structure. Therefore, much more attention in evaluation research should be given to the relationship between the tendering and market conditions on the one hand and the behaviour of both providers and job seekers on the other.

List of abbreviations

CWI	Centres for work and income
DEWR	Department of Employment and Workplace Relations
DEWRSB	Department for Employment, Workplace Relations and Small Business
RWI	Council for Work and Income [<i>Raad voor Werk en Inkomen</i>]
SUWI	Structure of Work and Income Implementation [<i>Structuur Uitvoering Werk en Inkomen</i>]
UWV	<i>Uitvoering Werknemersverzekeringen</i>
WAO	<i>Wet op de arbeidsongeschiktheidsverzekering</i>

References

- ANAO – Australian National Audit Office. *Management of the implementation of the new employment services market*. Canberra: ANAO, 1998 (Audit Report No 7, Performance Audit).
- Bartlett, W. et al. *A revolution in social policy. Quasi-market reforms in the 1990s*. Bristol: The Policy Press, 1998.
- Bouckaert, G. Sustainable development of networks in a governance context. In: Theron, F. et al. *Spanning the global divide: networking for sustained*. Stellenbosch: School of Public Management, 1998.
- Canoy, M.; Vollaard, B. Middelen om te bemiddelen. *Economisch Statistische Berichten*, 1999, No 4213, p. 536-538.
- Clarke, A. *Evaluation research. An introduction to principles, methods and practice*. London: Sage Publications, 1999.
- Cordia, L. et al. *De kunst van het aanbesteden. Onderzoek naar werking van de aanbestedingsprocedure bij inkoop van reïntegratieactiviteiten*. Zoetermeer: College van Toezicht Sociale Verzekeringen, 2001.
- Davidson, J. *Unpacking the shape of the new employment services market*. Jobs Australia 1998 National Seminar. Melbourne, 1998.
- Deakin, N.; Walch, K. The enabling state: the role of markets and contracts. *Public Administration*, 1996, No 1, p. 33-48.
- DEET – Department of Employment, Education and Training. *The SkillShare Network: an evaluation*. Canberra: DEET, 1993 (EMB Report 5/93).
- Derksen, W. et al. *Het borgen van publiek belang*. The Hague: Sdu Uitgevers, 2000.
- Descy, P.; Tessaring, M. (eds). *Training and learning for competence: second report on vocational training research in Europe. Synthesis report*. Luxembourg: Office for Official Publications of the European Communities, 2001 (Cedefop Reference series, 6).
- DEWR. *Employment Services, an active participation model*. Canberra: DEWR, 2002a (Discussion Paper).
- DEWR. *Government response to the Productivity Commission Independent Review of Job Network*. Canberra: DEWR, 2002b.
- DEWR. *Job Network evaluation, Stage Three: effectiveness report*. Canberra: DEWR, 2002c.
- DEWRSB. *Job Network evaluation. Stage One: implementation and market development*. Canberra: DEWRSB, 2000a.
- DEWRSB. *Job Network evaluation. Stage Two: progress report*. Canberra: DEWRSB, 2000b.
- DEWRSB. *Job Network: a net impact study*. Canberra: DEWRSB, 2001 (EPPB Report No 1/2001).
- Dijk, S. et al. *Een markt in ontwikkeling? Eerste inventarisatie van de stand van zaken op de markt voor reïntegratiediensten*. The Hague: Elsevier, 2000.
- Domberger, S. *The contracting organisation – a strategic guide to outsourcing*. Oxford: Oxford University Press, 1998.
- Eerste Kamer. *Wet Structuur Uitvoeringsorganisatie Werk en Inkomen*. Memorie van Antwoord, 9 oktober 2001, Eerste Kamer der Staten-Generaal Vergadering 2001-2002. The Hague: Eerste Kamer, 2001.
- European Commission. *Modernising Public Employment Services to support the European employment strategy*. Communication from the Commission. Luxembourg: Office for Official Publications of the European Communities, 1998 (COM(98) 641 final).
- Fay, R. G. *Making the Public Employment Service more effective through the introduction of market signals*. Paris: OECD, 1997 (Labour Market and Social Policy Occasional Papers, No 25).
- Fisher, C. M. *Resource allocation in the public sector. Values, priorities and markets in the management of public services*. London/New York: Routledge, 1998.
- Flynn, N. *Public sector management*. New York: Harvester Wheatsheaf, 1993.
- Green, A. et al. Financing vocational education and training. In: Descy, P.; Tessaring, M. (eds). *Training in Europe: second report on vocational training research in Europe 2000. Background report*. Luxembourg: Office for Official Publications of the European Communities, 2001, Vol. 1, p. 71-113 (Cedefop Reference series, 3008).

- Groot, I. et al. *Lessen uit de privatisering van reïntegratiediensten*. Amsterdam: SEO, University of Amsterdam, 2002.
- Hodge, G. A. *Contracting out government services: a review of international evidence*. Melbourne: Mony Tech Pty, 1996.
- Le Grand, J. From knight to knave? Public policy and market incentives. In: Taylor-Gooby, P. (ed.) *Trust and welfare*. Houndmills: Macmillan, 2000.
- 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 29.9.2003].
- Le Grand, J.; Bartlett, W. (eds). *Quasi-markets and social policy*. London: Macmillan Press, 1993.
- LISV. *Evaluatie aanbestedingsprocedure. Deel 2 eindrapport*. Amsterdam: Landelijk Instituut Sociale Verzekeringen, 2001.
- Mevissen, L. et al. *Op weg naar volwassenheid? Tweede inventarisatie van de stand van zaken op de markt van reïntegratiediensten*. Doetinchem: Reed Business Information bv, 2002.
- Näringsdepartementet. *En effektivare arbetsmarknadsutbildning*. Stockholm: Näringsdepartementet, 2000 (Ds 2000:038).
- OECD. *Innovations in labour market policies. The Australian way*. Paris: OECD, 2001.
- Parsons, W. *Public policy: An introduction to the theory and practice of policy analysis*. Cheltenham: Edward Elgar Publishing, 1995.
- Patton, M. Q. *Practical evaluation*. London: Sage Publications, 1982.
- Productivity Commission. *Independent review of the job network*. Canberra: Commonwealth of Australia, 2002 (Inquiry Report No 21).
- Reh fuss, J. A. *Contracting out in government: a guide to working with outside contractors to supply public services*. San Francisco: Jossey-Bass, 1989.
- Revenboer, A.; Lantinga, H. Lessen voor de reïntegratiemarkt. *Reïntegratie*, 2002, Vol. 2, No 6, p. 4-7.
- Savas, E. S. *Privatization and public-private partnerships*. New York: Chatham House Publishers, 2000.
- Struyven, L. et al. *Van aanbieden naar aanbesteden. Marktwerving bij arbeidsbemiddeling en -reïntegratie in Australië, Nederland, Verenigd Koninkrijk en Zweden*. Leuven: Acco, 2002.
- Struyven, L.; Steurs, G. *The competitive market for employment services in the Netherlands*. Paris: OECD, 2002 (Labour Market and Social Policy Occasional Papers).
- Struyven, L.; Steurs, G. Towards a quasi-market in reintegration services: first assessment of the Dutch experience. *Australian Journal of Labour Economics*, 2003, Vol. 6, No 2.
- SZW. *Naar een werkende reïntegratiemarkt*. Regeringsnota, 28 November 2002. The Hague: Ministerie van Sociale Zaken en Werkgelegenheid, 2002.
- Thompson, G. et al. *Markets, hierarchies and networks: the Coordination of social life*. London: Sage, 1991.
- Thuy, P.; Hansen, E.; Price, D. *The employment service in a changing labour market*. Geneva: International Labour Office, 2001.
- Trommel, W.; Van Heffen, O.; van der Veen, R. Marktwerving in de publieke sector. *Beleid en Maatschappij*, 2001, Vol. 28, No 3, p. 130-138.
- UWV. *Evaluatie aanbestedingsprocedure 2002*. UWV, 2002. Available from internet: <http://www.uwvnet.nl> [cited 30.9.2003].
- van der Veen, R. *Restructuring a corporatist welfare state. Managed liberalisation and the role of the state, the social partners and the market in Dutch social security and labour market policy*. Paper for the conference 'The transitional labour market. Outlines of an active labour market policy' held in Brussels, 2.3.2000.
- Walch, K. *Public services and market mechanisms. Competition, contracting and the new public management*. Basingstoke: Macmillan, 1995.
- Walch, K. et al. *Contracting for change: contracts in health, social care and other local government services*. Oxford: Oxford University Press, 1997.
- Webster, E.; Harding, G. *Outsourcing Public Employment Services: the Australian experience*. Melbourne: Melbourne Institute, 2000 (Working Paper, No 4/00).
- Williamson, O. *Markets and hierarchies: analysis and antitrust implications*. New York: The Free Press, 1975.
- Williamson, O. *The economic institutions of capitalism*. New York: The Free Press, 1985.



Evaluation activities in the European Commission

Josep Molsosa

Abstract

The purpose of this contribution is to review the current context and the prospects for evaluation activities within the European Commission. Under its President, Mr Prodi, the Commission has been radically overhauling its internal working methods and planning and decision-making procedures since 2000; this is having a significant impact on the role of evaluation within Commission services.

It is, therefore, of interest:

- (a) to give a brief overview of evaluation activities prior to the reform, especially in the areas of education and vocational training;*
- (b) to assess the Commission's new guidelines;*
- (c) to predict the changes that these guidelines are likely to bring about (objectives, methods, organisation, etc.) in evaluation activities.*

This contribution aims to encourage, promote and offer pointers for future research into the new culture of planning and evaluation of Community activities.

Table of contents

1. Brief overview	275
1.1. Policy and programme evaluation: some background	275
1.2. Review of evaluation activities	275
1.3. Community policies and programmes on education and training	276
1.4. Evaluation of Community education and training programmes	277
1.5. Evaluation results and evaluability problems	277
1.6. Limited use of evaluation results	279
2. Commission reform and new guidelines	281
2.1. Reform of strategic decision-making	281
2.2. Improving evaluation at the Commission	281
3. The new method challenges	283
3.1. Context and policy prospects in education and training	283
3.2. Evaluating education and training policies and programmes	284
List of abbreviations	286
References	287

1. Brief overview

1.1. Policy and programme evaluation: some background

Though not an exhaustive picture of evaluation activities at the Commission, some background is given below to provide a framework for the following analysis.

First is the reform of the Structural Funds in 1988. This reform marked a transition from an intervention approach based on support for a large number of projects (tens of thousands per year under the European Social Fund) to an approach involving planning and programming in consultation with the Member States. This new approach made it necessary to use evaluation practices for policies and programmes; the Commission services had had little experience of this at that time, as they tended to focus on evaluating 'small' projects.

Second are the efforts to clarify concepts. Methodological documents on the logical framework and on project cycle management, promoted by Directorate-General Aid and Cooperation in 1992-93, were of considerable use to those evaluating the three Structural Funds. An overall methodological document, *Common guide for monitoring and interim evaluation*, was published in 1995. In parallel, the MEANS (methods of evaluation of structural actions) initiative made it possible to consolidate evaluation methods and practices for this type of policy and programme. DG Agriculture also developed its own evaluation guides within the framework of the Common Agricultural Policy.

Third, the support and assistance activities of DG Budget, and in particular the organisation and leadership of an evaluation network within the Commission in order to exchange experience, methods and the results of internal practices. The 1997 document, *Evaluating EU expenditure programmes*, paved the way for a very useful systematisation of evaluation methods and practices.

The pioneering work of DG Budget in this area enabled the Commission to adopt a Communica-

tion on evaluation (European Commission, 2000b), to promote good practice (European Commission, 2000a), and to reform programme preparation methods (European Commission, 2001c). This was done in accordance with the Financial Regulation and pursuing the efforts to improve management methods that the Santer Commission had launched prior to the 1999 crisis.

1.2. Review of evaluation activities

In general, Commission services have made major progress in recent years in evaluating policies and programmes. The Communication on evaluation of 26 July 2000 states: 'reviewing the situation in 1995, the Commission concluded that, in a formal sense, the evaluation policy had been successfully implemented: an increasing number of evaluation reports are completed in most policy areas, an annual evaluation review is published, evaluation systems are regularly reviewed against agreed good practice guidelines, the Network of Evaluators is operative and evaluation reports are increasingly available to the public. However, it was acknowledged that further progress was needed, in particular as regards the quality and use of evaluation and its relevance for decision making.' (European Commission, 2000b).

Problems, nevertheless, continue to exist. The following are some of the conclusions, by way of example, from the *Annual evaluation review 2001*: 'there are some weak points in the Commission's policy instruments or management methods, revealed in several evaluation reports in different policy areas. However, it is not possible to draw wider conclusions without more detailed investigation of the background and reasons behind these observations. Nevertheless, it is significant that different evaluators in different policy areas are repeating similar types of criticism – some of it already in previous years' (European Commission, 2002a).

These are the main weak points:

- (a) the need to reduce the administrative burden imposed by the Commission on the beneficiaries or participants in EU programmes. The complexity of procedures had already been criticised in previous years. In 2001, for example, it was highlighted in the following evaluation reports: authorisation of medications, IDA II programme, 5th Framework programme monitoring report, IST programme, Youth for Europe, food aid and in some reports on external assistance and development aid, PHARE and ECHO programmes;
- (b) the need to reinforce the definition of clear strategy and clear objectives for Community interventions. A lack of focus can be a problem, even in interventions that are considered effective. This is the case, for example, in the support for forestry measures in agriculture and humanitarian aid policy. In development policy, there is a risk that NGOs 'dictate', policy when interventions are only 'demand driven';
- (c) shortcomings of monitoring systems and lack of appropriate data continue to be recurrent observations in evaluation reports, as noted in previous years. Evaluations strongly urge strengthening project and programme monitoring and introducing more structured monitoring procedures, while avoiding unnecessary bureaucracy.

1.3. Community policies and programmes on education and training

The current main programmes in these fields (Socrates and Leonardo da Vinci [LdV] in particular) are the outcome of a process of consolidation within the Commission, in conjunction with the Member States, of the initial programmes launched in the 1980s. These followed the formation of the Task force RHEFJ, which was to have greater operational autonomy in comparison with the highly financial approach taken by the European Social Fund at that time.

In vocational training, drawing on experience with small programmes or initiatives such as PETRA, Comett, Eurotecnet, FORCE, which

supported small pilot projects in innovation, exchanges and cooperation networks, mobility, etc., systematic thinking in the form of memoranda (on the concentration of actions in vocational training, for instance, as well as on rationalisation and integration in 1990-91) led to the 'integrated' programmes of 1995-99. It should be noted that the LdV programme was originally an action programme for implementing European Community vocational training policy.

In culture, this concentration took place in 1999 in order to rationalise and improve the cultural programmes Kaleidoscope, Ariane and Raphael, set up in 1996-97. The new programme Culture 2000 for the period 2000-04 was the result of a process of consultation and reflection as well as evaluation exercises.

The logic of these instruments was to promote and stimulate innovation and bottom-up exchanges, attempting to involve the largest possible number of participants/partners. It was also a way of consolidating and playing a part in the construction of Europe. Visibility was very important in this respect, allowing participants to feel that they were actually 'experiencing' Europe. Less important than impact on systems, the implicit objective was to encourage activism: the aim was to get things moving and to get the maximum number of people, and those responsible for education and training activities at the various levels, to work with one another.

The continuation of these programmes had been decided during 1998 and 1999 at a time when the Commission was subject to strong pressures following, among other things, the management problems of the LdV programme. With fairly formal *ex ante* evaluation exercises, the second phases of Socrates and LdV were a continuation of the intervention logic; this was despite some simplification and concentration of objectives and measures, thanks to a slight budget increase.

From an operational point of view, the new programmes included many more decentralised procedures and methods: in some cases this was not the best solution (see the networks in the context of LdV II). In evaluation, decisions made it compulsory, for the first time, to follow a set sequence: interim evaluations, proposal on possible continuation and *ex post* evaluations.

1.4. Evaluation of Community education and training programmes

Evaluation activities have covered action programmes, but not Community policies in these areas.

The main evaluation activities for education/training/youth programmes in recent years have included:

- (a) 1997-98
 - (i) interim evaluations of the Socrates I, LdV I and Youth for Europe programmes;
 - (ii) *ex ante* evaluations of the programmes following on from them;
 - (iii) evaluations of the Raphael, Kaleidoscope and Ariane programmes, and *ex ante* evaluation of the future framework programme Culture 2000;
- (b) 1999-2000: launch and finalisation of the final or *ex post* evaluations of programmes;
- (c) 2001: a year of transition without much programme evaluation, but during which the external evaluation of Cedefop was finalised;
- (d) 2002: a number of evaluations of actions or initiatives of DG Education and Culture (EAC):
 - (i) *ex ante* evaluations of the e-learning initiative and Erasmus-World;
 - (ii) operational evaluations of the NARIC network, the European year of languages and Eurydice.

Other thematic evaluations have also been launched, to assess specific aspects of the implementation of measures or actions, or to measure the impact of programmes on specific target groups or beneficiaries.

The main evaluation exercise, however, was the launch in late 2002 of work to prepare proposals for future programmes for the period 2007-13. Bearing in mind that 2004 and 2005 will be years of change (elections to the European Parliament, new Commission), which will have an impact on the ways in which new decisions are taken, DG EAC felt that this work should be started as soon as possible in order to be able to put forward proposals in late 2003/early 2004. Building on the results of all the preceding evaluation work, and using the new evaluation

approaches decided by the Commission, offer opportunities and pose challenges (see Section 3.2).

1.5. Evaluation results and evaluability problems

The following are some overall comments by way of review.

Ex ante evaluations are still seen too often as formal exercises that have to be carried out, alongside the memorandum or statement of grounds, for the financial statement required by the Financial Regulation. It is difficult to see them as genuinely independent evaluations, especially as they are generally written when the decisions have already been taken.

Interim evaluations, focusing on operational matters, are those which provide the most useful information for managers, as they can be used to improve implementing procedures or to draw up better targeted priorities. In general, however, they have not been used to assess the conceptual relevance and coherence of instruments, which should be one of the objectives of interim evaluations: finding out whether the approach being taken is the right one.

The purpose of *ex post* evaluations is, in practice, to report on and justify what has taken place. In many cases, they can be seen as reports on implementation rather than genuine evaluations.

In general, an overview of the methods of evaluation of Community programmes in education and training shows that ⁽¹⁾:

- (a) all evaluation reports provide a great deal of information on achievements, outputs (number of projects, number of participants, fields, financial data, etc.), but find it difficult to measure results against the specific objectives of programmes and, in particular, their impact on systems, which is, in general, the objective of such intervention;
- (b) all the reports detail implementation and follow-up problems: complex and cumbersome procedures, delays in selection and decision-making, etc.;

⁽¹⁾ Other analyses of the results of this kind of evaluation are contained in this report or can be consulted on the DG EAC website: http://europa.eu.int/comm/dgs/education_culture/index_en.htm [cited 15.4.2004].

- (c) most report a lack of complementarity with other Community policies and problems linking up with national and/or regional policies and initiatives in the same field;
- (d) the conclusion is that these are programmes which help promote innovation and exchange but which are not genuine instruments serving what should be a Community policy; etc.

It should be possible to improve some of these problems during the programme second phase, even though they have been negotiated in an undoubtedly difficult context. In practice, improvements include a reduction in objectives (even if this is only formal in some cases), an attempt to simplify measures and procedures (even if there are still too many in some cases) and decentralisation of some tasks to national agencies (which may limit Community follow-up of the results).

The Commission's services are making significant efforts to improve the management of these programmes and of outside technical assistance. It would seem, however, that this improvement work is still meeting limitations, and that there is still a long way to go to provide operational instruments that can be genuinely controlled and that serve Community policies.

An initial explanation/justification of this conclusion has to be sought in the characteristics of programmes themselves, which are the outcome of a difficult process of integration of old practices or priorities and are, in some cases, designed in an intentionally ambiguous way in order to allow participation by any project promoter. Priority is given to number and diversity of direct beneficiaries rather than to any attempt to make these programmes into an instrument of active policy. These choices are probably explained by the process of building the Community.

To understand all these problems, the features of these programmes need to be analysed in further detail. This does not just entail an analysis of formal decisions, but an analysis of actual practices, choices of annual priorities, selection criteria for projects and promoters, etc.

It is when evaluations are being prepared, when questions are being drawn up, when the actual logic of the action measure is being

explained, and when the theory on which it is based is being explained as well as the way in which it is supposed to achieve the objective, that these problems are encountered ⁽²⁾.

When it is attempted to transform the general, often vague, questions that those commissioning an evaluation have in mind into questions that are precise enough to be able to be handled using evaluation research methods (invariably based on research methods from economic, social, management and other sciences), the practical result, as pointed out by DG Budget's evaluation manual (European Commission, 1997), is that, 'retrieving the original intervention logic of a programme is sometimes easier said than done. Official documents often do not contain any systematic description of causal assumptions. Even the programme's objectives may only be stated in a very limited fashion. In any event, when a programme's objectives are not given a sufficiently transparent and precise meaning, it can be very difficult to judge its success'.

This raises the question of the evaluability of programmes. Evaluations need to be designed in the most realistic way. It is always useful to evaluate a programme, even if it is necessary to confine the evaluation to questions for which answers can be provided with a minimum of rigour and reliability.

The relative ambiguity between the general objectives set out in decisions and the content of the concrete measures included in programmes to achieve them has to be seen in this context. It is probably in the actual design of programmes that it is possible to locate the causes that limit their efficiency in terms of impact and the possibilities of measuring this impact.

In more concrete terms, and by way of pointers for analysis:

- (a) as regards implicit objectives: the aim is to reach a large number of beneficiaries, promoters, participants, etc., but also to obtain high-quality results with a clear-cut Community added value and a significant impact on systems. Are the measures, as designed, able to contribute to this?
- (b) as regards the steering of instruments to serve Community policies: the aim is to step up the Commission's capacity to draw

⁽²⁾ See the latest interim evaluation of the LdV programme (September 2003, accessible on the Europa site).

conclusions and make proposals to the Member States, but procedures which are in some cases too decentralised (the networks in LdV for instance), or too complicated, make it impossible to work towards conclusions or proposals that can be used for policy-making purposes in a timely way;

- (c) as regards operational aspects: the aim is to work with many projects, with promoters able successfully to manage a process of innovation, formalise their results and disseminate them widely among other partners and the Community and national authorities concerned. These are, nevertheless, very small projects with an annual budget which is still very limited in comparison with requirements, and they are managed by very committed promoters who lack the capacity to exploit and disseminate results;
- (d) as regards management and follow-up, the aim is to promote and follow up concrete achievements and the results progressively obtained but people are inclined to monitor financial aspects which, although necessary, do not in themselves provide any real Community added value;
- (e) as regards complementarity with other Community policies, or with other national or regional policies, programmes generally contain simple statements of intent (of the type 'the Commission shall ensure complementarity with [...]'), which are not actually translated into measures or into the design of the programme.

In practice, programmes are the result of an ambiguous conception of the intervention logic, which is half-way between two logics or two separate approaches:

- (a) an approach geared towards promoting basic innovation among a large number of small actors, who cooperate with other small actors in other Member States, making it possible to disseminate the European idea. This is a 'sprinkling' approach which undoubtedly leads to highly visible positive results, but has no real role to play in shaping policy;
- (b) an approach geared towards supporting policies and practices of innovation within the education and training systems and practices of Member States. This is a 'laboratory' approach whose main beneficiaries are those

responsible for designing and implementing policies and decision-makers. It requires more specific objectives, more demanding criteria, measures that are better designed and linked with one another, and simpler and more controllable operating procedures.

It is probably this ambiguity that explains why these programmes are difficult to evaluate and why it is difficult for evaluators to produce rigorous, exhaustive and useful evaluations. A good number of evaluation reports are, in practice, implementation reports that 'evaluators' are asked to draw up, bearing in mind the problems that Commission services have in actually following up the implementation of programme measures.

1.6. Limited use of evaluation results

Despite the problems mentioned above, evaluators provide analyses, conclusions and proposals. These results can only be put to limited use, bearing in mind the following.

Timing is a concern. Judging by the reports and communications set out by decisions, and since programmes understandably cannot be made to 'stop in their tracks', successive phases of programmes have to be decided upon before the results of the evaluations of the previous stages are available. For instance, the second phase of the LdV programme (adopted in 1999 for the period 2000-06), was prepared from the interim report on implementation, drawn up in late 1997, which summarised the analyses of the implementation reports submitted by the Member States in late 1996.

The range of commissioners/beneficiaries of results of evaluations also offers problems. In vocational training, for instance, the range of guidance, cooperation and policy-making structures (Advisory committee on vocational training, LdV programme Committee, Employment and Education Councils, European Parliament Committees, Directors-General of vocational training, Cedefop's Management Board, etc.), national authorities, education agencies, etc., means that the steering of evaluation exercises (priority questions, terms of reference of evaluations), actual conduct (management of external contracts, acceptance of evaluators' reports, etc.) and the use of results, is complex.

Despite real progress, the new culture of evaluation introduced by the Commission has been assimilated only to a limited extent in some Commission services:

- (a) evaluation is still seen too often as a formal obligation to justify expenses, rather than as an effective way of preparing for new actions, or as an efficient way of improving the ways in which current actions are implemented;
- (b) senior staff and managers are overly occupied by direct management of programmes, especially by administrative and financial procedures. This leads, in some cases, to evaluation plans whose design, rigour and quality could probably be improved;
- (c) basic concepts are still not widely known and used by officials to prepare evaluation plans,

even the three simplest concepts set out in Commission documents ⁽³⁾. Despite training schemes, there is still work to be done to promote the correct use in practice of the conceptual framework 'resources → outputs → results → outcomes' within management units responsible for setting up evaluations.

The budget and deadlines available for many evaluations, which have to measure results and impact throughout Europe, in some cases limit the quality of the offers received. Specifications, which in some cases are overly ambitious in comparison with the financial resources available, discourage many experienced experts and lead to offers only from experts with limited experience of the requirements of Europe-wide evaluations.

⁽³⁾ *Defining objectives and progress indicators*, SPP Unit of the Secretariat-General, 2002.

2. Commission reform and new guidelines

2.1. Reform of strategic decision-making

Following the 1998/1999 crisis (the Commission resigned in March 1999, following the report by the Committee of Wise Men), the new Prodi Commission made working methods in the Commission one of its strategic priorities ⁽⁴⁾.

The proposals of the White Paper on reforming the Commission (March 2000), were justified by President Prodi, before the European Parliament, when he said that there was a need for a genuine revolution in the Commission's working methods: a policy planning mechanism, including the definition of policy priorities and activity-based management, should make it possible to ensure the necessary match between objectives, actions and resources.

As regards decision-making, the Commission has adopted an activity-based management approach (July 2001), whose main objectives are to:

- (a) increase cost awareness through integrated decisions on priorities, objectives, activities and the allocation of human, administrative and financial resources;
- (b) tie in strategic planning with the operational programming of activities and monitoring and evaluation of implementation,
- (c) develop performance-based management by emphasising results rather than input control.

These are the operational principles:

- (a) a new programming, decision-making, management and evaluation culture which is more systematic, more streamlined, more transparent and more efficient;
- (b) all decisions are to be taken from a systematic knowledge base, detailed and transparent analyses and rigorous evaluations;
- (c) evaluation becomes a management tool and a tool for planning and programming activities;
- (d) the annual planning and programming cycle comprises the following elements:

- (i) the annual policy strategy, decided by the Commission in consultation with its services;
- (ii) the annual management plans of each DG;
- (iii) the Commission's work programme;
- (iv) the annual reports.

2.2. Improving evaluation at the Commission

In the context of the reform of the Commission, the Communication of July 2000 on evaluation, mentioned above, proposed a number of routes or measures to improve evaluation activities:

- (a) refocus or supplement mandatory evaluations with other evaluations able to serve as a basis for informed decision-making in the annual strategic planning and programming cycle;
- (b) gear evaluations towards more political and more strategic issues;
- (c) ensure that evaluations are able to provide results in a timely way;
- (d) adapt methods, the work plan and the timetable to ensure that evaluations make an effective contribution to decision-making;
- (e) make *ex ante* evaluations systematic and improve them (needs analysis, setting of objectives, alternative forms of intervention, checking of Community added value, cost-benefit analyses, etc.);
- (f) ensure that monitoring systems are provided with information on inputs, outputs and the results of activities.

In order to ensure the correct preparation of new proposals, the Commission has decided to require prior evaluations of impact, to accompany the proposals of the Directorates-General, as part of preparations for the annual policy strategy and the annual work programme.

⁽⁴⁾ Reference documents may be obtained, in all languages, from the Commission's website: <http://europa.eu.int/comm> [cited 15.4.2004].

This requirement is the key to effective integration of evaluation into the annual programming cycle. It is explained in the *Communication from the Commission on impact assessment* (European Commission, 2002c).

In practice, it is a prior evaluation of the estimated impact of the various possible options and instruments. It is a method that aims to structure the process of policy-making. It is the systematic analysis of:

- (a) the problem addressed by the proposal concerned;
- (b) the objective it pursues;
- (c) the alternative options available to reach the objective;
- (d) their likely impacts;
- (e) respective advantages and disadvantages, including synergies and trade-offs.

The impact assessment is not a fiche that is filled in just before the Commission adopts a proposal. It is not a checklist with boxes to tick. It requires thinking about the proposal's possible impacts and how to assess them from the beginning of the process and continuing to review

them until its final elaboration. Both the impact assessment and the *ex ante* evaluation have to be seen as a process that supports the preparation of proposals for new or renewed Community actions.

According to the Commission document, the purpose of impact assessment is to improve the quality and coherence of policy formulation. Impact assessment is an aid to decision-making and not a substitute for political judgement. In practice, political judgement involves complex considerations which go far beyond the anticipated impacts of a proposal. An impact assessment does not necessarily generate clear-cut and decisive conclusions and recommendations. It nevertheless makes an important contribution by informing decision-makers of the consequences of policy choices.

This communication is supplemented by a further *Communication on evaluation standards and good practice* (European Commission, 2002d), which is to be progressively implemented by services to ensure that evaluations are high-quality and useful and are put to practical use.

3. The new method challenges

What changes are these new guidelines and the major overhaul of the methods of designing and deciding on Community actions likely to bring about in evaluation methods and practices?

Despite the problems and the apparent overload that the first cycles entailed, the Commission reform is becoming established in people's minds and practices, and the new management culture is gradually taking root. The integrated approach to strategic planning includes the evaluation function in an intelligent way.

This new trend offers an opportunity for evaluators to make their work more useful. It requires, however, a major change in approach. There needs to be a shift towards forecasting, which may actually make decision-making easier.

Use needs to be made of methods able to provide ideas and useful information in a timely way, to assist decision-making, which is increasingly complex and based on an increasingly systematic knowledge base; it is also grounded in economic, social, technological, cultural, demographic and other environments where change is also gathering pace.

New methods need to be found in which rigour, coherence and the relevance of analyses are compatible with deadlines and available resources, in order to design and adopt effective solutions for the new and expanding issues facing Europe.

From a more practical point of view, there needs to be a shift towards merging ex ante and impact assessment methods. The Commission, which has so far drawn up a whole range of evaluation and monitoring methods, should finally provide itself with methodological base for programming its activities.

3.1. Context and policy prospects in education and training

Before looking at ways of promoting or guiding prospective research on the evaluation of policies and programmes in education and training, some

aspects of the context and of prospects for these policies need to be examined in more detail.

First is the gradual 'politicisation' of DG EAC initiatives and the launch of new strategy proposals. Action programmes (Socrates, LdV) are gradually being supplemented and framed by policy thinking and proposals. These include the Memorandum and Communication on lifelong learning, the White Paper on Youth, the promotion of a European area of knowledge, open coordination at Community level of the objectives of education systems, etc. Other initiatives are being prepared on mobility, accreditation, transparency and the recognition of qualifications.

The Bologna and the Bruges processes ⁽⁵⁾ are indicators of this increased push for cooperation and policy coordination and received significant support at the Lisbon European Council in March 2000 ⁽⁶⁾.

Enlargement to a Europe of 25 Member States will undoubtedly influence the content of policy proposals, for instance towards systems innovation/restructuring and the consolidation of cooperation and exchange procedures between those responsible for national/regional education and training policies.

The Convention on the Future of Europe, set in motion by the Laeken European Council (December 2001), in its proposals on the EU's mission, confirms the Commission has a role to play in education and training. The Convention proposes that the EU should implement support measures in employment, education and vocational training, youth culture, sport, etc.

The adoption of the open method of coordination ⁽⁷⁾ seems to reflect the will of the Member States to step up such cooperation and exchanges in the form of 'open' coordination of the general problems in some fields, with two main objectives:

- (a) disseminating best practice;
- (b) ensuring wider convergence on the main objectives of the EU.

⁽⁵⁾ See information on the Commission website, DG EAC pages on political cooperation.

⁽⁶⁾ See conclusions of the Presidency (European Council, 2002).

This requires the Commission to demonstrate a better capacity to lead, analyse, evaluate and propose new and innovative strategies to improve the efficacy and efficiency of systems.

The second contextual aspect is the progressive integration of education and training, as shown by thinking about lifelong learning, the proposals that followed on from it, and the adaptation of the actual priorities of programmes to these proposals.

The third aspect is the process of preparing future programmes of the Commission in these fields, which started in September 2002. These future instruments will have to meet new political ambitions, the new requirements of an enlarged Europe and the new social, technological, cultural, economic and other issues that our societies and political and administrative systems are facing.

DG EAC will need instruments and programmes able to achieve more ambitious general objectives, instruments with well-defined and appropriately articulated specific objectives, accompanied by intervention mechanisms that are better adapted, more flexible and more ambitious; programmes that are, ultimately, clearer and easier to monitor and evaluate.

3.2. Evaluating education and training policies and programmes

The following are the main areas for consideration in improving evaluation practices (purposes, methods, organisation, etc.) within DG EAC, which require a contribution from specialists and researchers on evaluation methods and practices:

In the short term, there is preparation of, and concerted action on, new education and training programmes for the period after 2006. This offers a good opportunity to monitor approaches by the Commission to the preparation of proposals. What is needed in practice is the organisation of a policy-making process which draws on the results of the evaluations conducted up to now and which is able to offer decision-makers a

systematic information base, enhancing a decision-making culture based on analysis, instead of deal-making.

There needs to be a shift away from practices which sap the forward evaluation process, from the prior analysis of the impact of different options to the *ex post* drafting of an *ex ante* evaluation report justifying options that have been decided in an intuitive way.

Normally an internal process, this type of evaluation exercise requires clearer working methods and the sharing of responsibilities between managers/designers and those responsible for evaluations. The evaluation manual of DG EAC ⁽⁷⁾ sets out an iterative and parallel process in which evaluators work together with designers to ensure the rigour and quality of analyses and proposals.

Then come methods and approaches. Forward evaluations (*ex ante*, impact assessment) probably need to be improved to supplement the 'conventional' evaluations of decisions already taken (interim, *ex post*).

In terms of methods, and in order to ensure useful answers to relevant questions in a timely way, attempts should be made to set up a knowledge base, a systematically assessed record of experiences, that can be used at will and is continually updated to take account of the practices and experiences of all European participants.

The results of a continuing process of analysis would need to be fed into this database or expertise base, enabling integration of the diversity of Community experiences and participation by a large number of evaluation experts (in consultation exercises, seminars and workshops) appropriately geared to this objective of building on and making the most of expertise and evaluated experiences.

In practice, it will be necessary to place the critical observation of practices and methods on a systematic footing, to generate a capacity for analysis, evaluation and continuing exploitation. This should make it possible to exploit the results of evaluations as they become available and to ensure that European diversity is adequately represented, without being overly conditioned by the timing of the instruments used.

⁽⁷⁾ See conclusions of the Presidency (European Council, 2002, point 37).

⁽⁸⁾ Evaluation Manual. Basic concepts for the preparation, conduct and use of the results of evaluations. February 2003.

This approach will probably help managers/decision-makers to reduce the time needed to obtain useful and timely analyses. It will probably reduce the costs of analysis and evaluation work, and will ensure better quality contributions that are more objective and independent (European Commission, 2002d).

Finally, employing these methods, these forward and systematic approaches, will require thinking about resources and about the means and capacities needed.

In the first place, the Commission should study the possibilities for rationalisation, integration or coordination of existing capacities at Community level. In the new Union, how can the capacities of networks or agencies such as Eurydice, Cedefop or the European Training Foundation in Turin be better used?

In the Member States, what steps need to be taken to consolidate a genuine network of

experts and/or bodies that observe, analyse and evaluate education and training policies, particularly in a Community context in keeping with the policy priorities decided by the Commission in conjunction with the Member States?

At a more concrete operational level, what steps need to be taken to improve the quality and range of the expertise available in the various Member States? To make a start, DG EAC services could and should promote this new approach, this new culture of evaluation, among experts, especially those who have expressed interest in taking part in the evaluation work set in motion by the Commission (at the time of calls for expression of interest).

These are routes and initiatives to encourage thinking among researchers and to help to improve the design of Community policies, in order to ensure that they are effective and efficient.

List of abbreviations

DG	Directorat General
EAC	Education and Culture
LdV	Leonardo da Vinci

References

- European Commission – Directorate-General Budget. *Evaluating EU expenditure programmes: a guide. Ex post and intermediate evaluation*. Brussels: European Commission, 1997. Available from Internet: http://europa.eu.int/comm/budget/evaluation/pdf/guide_en.pdf [cited 15.4.2004].
- European Commission. *Good practice guidelines for the management of the evaluation function*. Brussels: European Commission, 2000a. Available from Internet: http://europa.eu.int/comm/budget/evaluation/goodpractice/goodpractice_en.htm [cited 15.4.2004].
- European Commission. *Focus on results: strengthening evaluation of Commission activities*. Communication to the Commission from Mrs Schreyer, in agreement with Mr Kinnock and the President. Brussels: European Commission, 2000b (SEC(2000)1051). Available from Internet: http://europa.eu.int/comm/budget/evaluation/pdf/sec20001051_en.pdf [cited 15.4.2004].
- European Commission. *Ex ante evaluation: a practical guide for preparing proposals for expenditure programmes*. Brussels: European Commission, 2001. Available from Internet: http://europa.eu.int/comm/budget/evaluation/pdf/ex_ante_guide_en.pdf [cited 15.4.2004].
- European Commission. *Annual evaluation review 2001 – Overview of the Commission's evaluation activities and main evaluation findings*. Working document of the Commission services. Brussels: European Commission, 2002a (SEC(2002) 337). Available from Internet: http://europa.eu.int/comm/budget/evaluation/pdf/review2001_en.pdf [cited 15.4.2004].
- European Commission. *Evaluation standards and good practice*. Communication to the Commission from the President and Mrs Schreyer. Luxembourg: Office for Official Publications of the European Communities, 2002b (COM(2002)5267). Available from Internet: http://europa.eu.int/comm/budget/evaluation/pdf/C_2002_5267_final_en.pdf [cited 15.4.2004].
- European Commission. *Communication from the Commission on impact assessment*. Luxembourg: Office for Official Publications of the European Communities, 2002c (COM(2002)276). Available from Internet: http://europa.eu.int/eurlex/en/com/cnc/2002/com_2002_0276en01.pdf [cited 15.4.2004].
- European Commission. *Communication from the Commission on the collection and use of expertise by the Commission. Principles and guidelines. Improving the knowledge base for better policies*. Luxembourg: Office for Official Publications of the European Communities, 2002d (COM(2002)713 final). Available from Internet: http://europa.eu.int/comm/governance/docs/comm_expertise_en.pdf [cited 16.4.2004].
- European Council. *Presidency Conclusions – Lisbon European Council: 23 and 24 March 2000*. Available from Internet: http://europa.eu.int/ISPO/docs/services/docs/2000/jan-march/doc_00_8_en.html#TOP [cited 16.4.2004].

List of authors

Mike Coles

Qualifications and Curriculum Authority
UK-London W1J 8QA
E-mail: ColesM@qca.org.uk

Loek Nieuwenhuis

STOAS Research
NL-6700 AB Wageningen
E-mail: Lni@stoas.nl

Hanne Shapiro

Danish Technological Institute
Centre for Competence and IT
DK-8000 Århus C
E-mail: hanne.shapiro@teknologisk.dk

Uwe Lauterbach

Hermann-Günter Hesse
Wolfgang Hellwig
German Institute for International Educational
Research
D-60486 Frankfurt am Main
E-mail: Lauterbach.Uwe@gmx.de

Sabine Fabriz

Institut für Pädagogische Psychologie
D-60054 Frankfurt am Main
E-mail: s.fabriz@paed.psych.uni-frankfurt.de

Bernd Baumgartl

Gerhard Schaumberger
Navreme Knowledge Development
A-1050 Vienna
E-mail: Baumgartl@navreme.net

Olga Strietska-Ilina

Cedefop
GR-Thessaloniki
E-mail: ost@cedefop.eu.int

Ludo Struyven

Higher Institute of Labour Studies
B-3000 Leuven
E-mail: ludo.struyven@hiva.kuleuven.ac.be

Geert Steurs

IDEA Consult
B-1000 Brussels
E-mail: geert.steurs@ideaconsult.be

Josep Molsosa

European Commission
B-1049 Brussels
E-mail: Josep.Molsosa@cec.eu.int

Cedefop (European Centre for the Development of Vocational Training)

Evaluation of system and programmes

Third report on vocational training research in Europe: background report

Pascaline Descy, Manfred Tessaring (eds)

Luxembourg: Office for Official Publications of the European Communities

2004 – 288 pp. – 21.0 x 29.7 cm

(Cedefop Reference series; 57 – ISSN 1608-7089)

ISBN 92-896-0279-1

Price (excluding VAT) in Luxembourg: EUR 25

No of publication: 3041 EN

This publication gathers contributions from renowned researchers on recent evaluations of education and training systems and reforms. The case studies provide insight into current evaluation practices in several European countries and assess the results of international and EU-supported programmes and initiatives.

This publication is part of the background report to Cedefop's third research report, dedicated to evaluation and impact of education and training. The other volumes address the *Impact of education and training* and the *Foundations of evaluation and impact research*. The third research report is complemented by a synthesis report which provides a comprehensive review of types and results of evaluation and impact research in Europe and beyond.

The series of Cedefop reports on vocational education and training research have been published since 1998. They take into account the wider context in which vocational education and training is embedded and discuss the implications for policy, practice and future research.

Pascaline Descy, Manfred Tessaring

EN

Evaluation of systems and programmes

Third report on vocational training research in Europe Background 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 25

3041 EN



Publications Office
Publications.eu.int

ISBN 92-896-0279-1



9 789289 602792 >