

Chapter 5

Public Employment Services: Managing Performance

How can the Public Employment Service (PES) assess the impact of its labour market programmes and use this information to manage them better? PES data systems need to allow identification of the “output” of labour market programmes in terms of their impact on off-benefit, employment and earnings outcomes. Impacts should be valued using the formula $(B + tW)$ where B is benefit payments saved, t is the tax rate on earnings and W is total earnings, i.e. the product of months employed and the monthly wage rate, with these outcomes measured for up to five years after the start of programme participation. In quasi-market systems, employment service providers must be given broad-ranging responsibility for clearly-defined groups of clients, and institutional arrangements must prevent “gaming” (artificial manipulation of outcome measures) and “creaming” (provider failure to enrol disadvantaged clients) and must protect individual entitlement to benefits. These underlying principles can be adapted to manage performance in traditional PES arrangements. Outcome measurement and the evaluation of programme impacts may seem to be relatively technical concerns, but they have already played an important role in the history of labour market policy in several OECD countries.

Introduction

Chapter 4 examined the impact of active labour market programmes (ALMPs). This chapter asks how the Public Employment Service (PES) can assess the impact of its labour market programmes and use this information to manage them better. In general terms, PES institutions and data systems need to allow identification of the “output” of labour market programmes, in terms of reducing unemployment and increasing employment and earnings, and use this information to replace less effective programmes with more effective ones. The chapter sets out preconditions for successful market-driven provision of publicly-financed employment services. These preconditions are often also relevant, although they may be relaxed or adapted in some respects, for the performance management of public services.

Section 1 surveys historical evidence that impact evaluation can be a driving force in the management of the PES and the results it obtains. Section 2 sets out some general principles for performance management. Section 3 considers i) quasi-market arrangements where the government defines output measures and financing conditions for the delivery of public employment services by competing independent organisations, and ii) the application of performance management principles within a more traditional PES organisation.

Main findings

- **The governance structure for employment services is a major determinant of success.** For example, the PES must manage the referral of jobseekers to external labour market programmes so that the PES can measure “motivation” effects that arise before clients enter programmes and the employment outcomes that arise after exit from programmes. External service providers need to have broad-ranging responsibility for services delivered to clearly-defined groups of clients, so that the impact of their services on client outcomes can be reliably measured.
- **Labour market authorities should track off-benefit, employment and earnings outcomes for programme participants for about five years.** PES management often takes benefit caseload decline or short-term post-programme employment rates as a measure of success, because these are the most visible and easily-measured outcomes. However, it is also important to assess which programmes have genuinely beneficial long-term impact.
- **Outcomes can be assessed in terms of a “B + tW” formula.** To a first approximation, programmes should be evaluated in terms of their impact on $(B + tW)$, where B is the benefit payments saved, t is the tax rate and W is total participant earnings (the product of employment rate and wage rate). When impacts are measured over long periods, the earnings component in this formula can be relatively large. Effective performance management with outcomes valued according to the $(B + tW)$ formula would not only reduce total unemployment but also increase the delivery of substantive employment

services which improve long-term employment and earnings outcomes. It would improve government's net financial balance, because the $(B + tW)$ criterion means that programmes are selected when the benefit savings and increased tax receipts that they generate exceed their cost.

- **Measures of outcomes and impacts must be hard to “game”.** When employment services are subcontracted, government agencies should assess outcomes from employment services in terms of the number of their clients who remain on benefits and/or who are in employment according to official data sources, not data reported by the service providers themselves. Countries could consider using tax and social security data records to track employment and earnings outcomes at low cost, subject to arrangements to prevent access to individual-level data.
- **“Creaming” (i.e. selection by service providers of which clients to serve) should be prevented.** Government should manage referrals to service providers and ensure that employment outcomes are measured for *all* persons referred to a provider. The measurement of employment outcomes this way creates no incentive for providers to divert their less-easily-employable clients to other service providers or to other welfare benefits.
- **Government should protect individual entitlement to benefit.** Service providers need to be able to report evidence of lack of availability for work or refusal to participate in a labour market programme, but at the same time government needs to ensure that valid benefit entitlements are protected.
- **Providers or services that have little impact on jobseeker outcomes should be systematically reformed and where necessary replaced.** Although this is an obvious recommendation, it may be difficult to implement in practice because in centralised systems staff resist restructuring and in decentralised systems the actors that currently receive financing tend to oppose change.
- **This framework is broadly applicable.** The broad framework here is applicable to management of a quasi-market for employment services but also to the performance management of local employment offices within a public system. For individuals who have no benefit entitlement, or in developing countries where informal sector employment is widespread, the measurement of employment outcomes in terms of earnings that appear in tax or social security records would reward employment services for bringing their clients into formal employment.

1. Historical experiences with the use of impact evaluation for PES governance

It is generally not possible to observe results from ALMPs directly, which is why it is so important to carry out evaluations of programme impacts. For most other services, approximate but direct measures of output exist. If garbage collection is not done, householders complain. If highway maintenance work is not done properly, contract supervisors notice. But if an ALMP has no impact on outcomes, it has for practical purposes no output and yet this may not be known to any of the actors involved. So a special effort is needed to assess the impact of ALMPs, or to structure PES operations so that impacts on outcomes are rewarded directly.

Despite the potential technical difficulties of using programme evaluations to manage the PES, it can be argued that outcome measurement and impact evaluation have played

an important role in the history of labour market policy in several countries. This subsection briefly summarises this history.

At least since 1986, employment policy in the **United Kingdom** has been managed with close attention to comparative tracking of numbers registered unemployed as a function of the different interventions and services delivered. When the idea of conducting “Restart” interviews with all long-term unemployed was devised late in 1985, the Treasury was only willing to agree to pilot implementations which had to be evaluated before any nationwide extension. Historically, this was a turning point.¹ In 1987, a new organisation (called the Employment Service) was created. From the start this organisation conducted focused evaluations of its operations. In 1990, the Employment Service was given the status of an autonomous agency with an Annual Performance Agreement which defined multiple quantitative targets, starting with the target of 1.65 million placings of unemployed people in 1990/91 (Price, 2000).

A focus upon quantitative evaluation continued through the 1990s. According to the compendium compiled by Greenberg and Shroder (2004), the UK Employment Service undertook about half of all European “social experiments” in the area of labour market policies over this period. This drive in favour of experimentation and evaluation continues with, for example, recent evaluations of the New Deals, Employment Zones for the unemployed and programmes such as “work-focused interviews” for recipients of other benefits.

Occasionally, detailed programmes are implemented nationwide before impact evaluations have provided any clear result or despite existing evidence of limited impact. Some changes, such as computer systems and the 1996 overhaul of benefit legislation, by their nature cannot be tested in advance. Nevertheless, evaluation has been a critical principle behind the long series of operational changes that have helped to reduce registered unemployment from over 3 million in 1986 to well below 1 million now.

In 1981, legislation allowed states in the **United States** to experiment with requiring work in return for welfare. Using random assignment methods, the Manpower Demonstration Research Corporation carefully evaluated 11 of the experiments, and its 1986 report on programmes with strong work requirements found that these “make a difference. They increase employment and earnings of recipients and they reduce welfare dependency”. The apparent success of these programmes led to the passage of the Family Support Act of 1988 which established the Job Opportunities and Basic Skills training programme, requiring not merely registration but the participation of welfare mothers in work activities. In the early 1990s, many states were allowed to experiment with their welfare programmes. Soon, a sharp decline in welfare caseloads was under way in some states. This experience helped to structure the federal welfare reform legislation of 1996 (citation from Zellman *et al.*, 1999; see also Council of Economic Advisers, 1997). Subsequently, welfare reform has on the whole reduced welfare caseloads more than expected and not generated as much hardship as critics feared.

In May 1998, most public employment services in **Australia** were replaced by the Job Network, which in the first contract period delivered services through about 300 contracted organisations (OECD, 2001b). Although Job Network providers were rewarded for the number of employment outcomes, lasting at least three months, achieved by their clients (with an additional bonus for outcomes lasting six months), this incentive mechanism in itself had limited impact because the majority of provider income was derived from fees for “commencements” (i.e. initial registration of jobseekers with the provider) rather than

placements. In the short term (i.e. for the duration of the 1998-2000 contract), a strategy of commencing jobseekers but spending relatively little on further provision of services could be profitable.

In the first (1997) tender round of the new system, contracts were issued to a wide variety of providers, resulting in high variability of performance. Providers first received general (unpublished) advice about their individual placement performance early in 1999. The first “star ratings” with regression adjustments (to reflect differences in jobseeker characteristics and local labour market conditions faced by individual providers, so as to measure impacts rather than gross outcomes) were published in March 2001. These developments progressively made it possible for providers to identify their own good or bad performance and to change strategy, and for government to select providers on the basis of performance. DEWR (2003) reported an increase in the net impact of Job Network services, which it attributed to “the market developing, and in recent times, the introduction of the ‘Star Ratings’ system which has driven substantial performance improvements”. Extensive reforms were announced in 2002 and became operational in the third Job Network contract period which started in July 2003. Overall, recent improvements in both programme impacts and aggregate outcomes (described in Chapter 4) reflect both extensive research that has informed the general strategy, and the increased accuracy and influence of explicit measures of comparative provider performance.

Radical changes of labour market policy in **New Zealand** were implemented in 1998 with the integration of benefit administration and employment services into a single agency, Work and Income, and the introduction of internal targets for placements into stable work (defined as work lasting more than three months). Currently, each client is classified by employment counsellor, local office, region, etc. and performance in terms of numbers of stable employment outcomes is monitored at each of these levels. This reform did not go smoothly at first and in 1998/99, placements were far below target.² However, by 2000/01 placements had doubled to well above target (Wallis, 2001) and total unemployment began to decline by about 10% per year. The new organisation’s “can-do” philosophy had an increasingly positive influence as the results-oriented management approach stabilized, and the decline in unemployment accelerated with the introduction of a package of activation programmes in 2003 (see Chapter 4, Chart 4.1).

An improvement in the volume and accuracy of evaluations may have been an important background factor for the implementation of effective policies. Although the New Zealand authorities have for at least a decade published impact evaluations of specific labour market programmes, “employment programme evaluation has made considerable progress since 1998, when the evaluations were last reviewed... [a review in 1998] identified the need to develop consistent outcome measures, predefined success criteria for programmes and robust measures of the cost or cost effectiveness of programmes... These were all necessary precursors to being able to generate comparative information” (Johri et al., 2004). Dixon (2002) discussed the use of administrative data for evaluation purposes, and Maré (2002) presented estimates for the impact of a range of employment policy interventions.

In 1994, **Denmark** adopted a reform combining active policies with administrative programmes to ensure the implementation of benefit criteria. The reform was not rooted in a culture of programme impact evaluation (it reflected a consensus within the administration and among the social partners in favour of a reform to tackle unemployment, without

much evidence of doubts about the method to use). However, the case that benefit eligibility criteria are important and that the reforms had a substantial impact was advanced through research by the Ministries of Finance and of Labour in the late 1990s (see OECD, 2000, Chapter 4; and OECD, 2002, Chapter 4 for references). More recently, the policy of systematic referral to labour market programmes during the active period of benefits has been relaxed in favour of lighter interventions, still of an activation nature.³

In the **Netherlands**, activation principles were implemented through public policy in the 1990s (see OECD, 2003a, Table 4.3). Since 2000, key employment services have been delivered by private providers, which have clients referred to them through a public gateway, as in Australia. However, in contrast to Australia, central government does not manage contracts with providers: contracts for unemployment assistance beneficiaries are managed by municipalities.⁴ Related to this, there is little government evaluation of the functioning of the quasi-market arrangements and no national statistical framework for generating comparative ratings of provider impacts on outcomes.⁵

In **Switzerland**, a system for rating of local employment office performance in terms of off-benefit outcomes was implemented in 2000. The publication of these ratings was preceded by detailed research into the determinants of local office placement effectiveness (see OECD, 2001). The ratings helped to improve local employment office performance, driving the registered unemployment rate down. However, cantons with low unemployment rates queried the validity of the ratings and the linking of cantonal funding to the ratings (*bonus-malus*) was suppressed in January 2003 (OECD, 2004b).⁶

In sum, a management culture where outcomes are tracked, programme impacts estimated and less effective programmes are replaced with more effective ones, has historically been a key factor in the development of effective policies in the United Kingdom and the United States since the 1980s and in Australia and New Zealand recently. The Netherlands and Switzerland have developed systems for performance management, but in these countries a division of responsibility between national and local governments makes it relatively difficult to measure impacts achieved by employment services and implement changes on this basis.

2. General principles for performance management

This section sets out further details of how performance management principles should be implemented, highlighting some of the issues and constraints that apply in the field of labour market policies.

A. If the PES is decentralised, funding should be subject to performance management

Unemployment benefits and employment services are often *financed* at national (or some cases, regional, state or provincial) level.⁷ However, employment services are actually *implemented* at local level. The national level needs to maintain some control over the local level, because it finances employment services to help limit spending on nationally-financed benefits and employment services at local level do not necessarily see a strong incentive to enforce eligibility criteria against individuals who are voluntarily unemployed. Indeed, local communities may find it advantageous to have national government pay income support even to individuals whose availability for work is limited. The national level also needs to impose consistent reporting standards and ensure that expertise

gathered at national level is used to improve practices at local level. The local level should benefit from national-level services such as standardized information technology, material for use in training courses and research findings about effective placement strategies, and should also be able to work autonomously to adapt its strategy to unique local conditions and to leave room for experimentation. So there are needs for hierarchical control, for local autonomy and for two-way communication. These needs might be met through a traditional PES with performance management or through a quasi-market (discussed further in Section 3).

B. The impact of active programmes should be evaluated by the PES

PES institutions often account directly for much of the substantive spending on ALMPs.⁸ They continue to have an important role in cases where employment services are subcontracted and when long-term ALMPs are delivered by non-PES institutions. Independent programme providers are not well placed to assess the overall impact of their own services and the PES needs to manage labour market programmes using the findings from some kind of impact evaluation.

The quantitative measurement of outcomes and programme impacts is a relatively technical activity, as compared to the historical tradition of PES activity and management concerns. However, the environment of active labour market policy is not what it was 40 years ago. Improvements in the level of social protection over the post-war period have increased benefit dependency and costs. Seven European countries now spend more than 1% of GDP on active programmes alone. Such high levels of spending justify great efforts to ensure that spending is well managed. The cost of information systems and research can be easily covered if they are effective. At the same time, improvements in information technology and technical expertise are tending to make more sophisticated evaluation and performance management strategies viable.

Although reported impact evaluations are often technical, this is partly due to a publication bias. Relatively simple evaluations by the PES have often given results which are relevant for operational purposes. For example, the pilot implementations of Restart in the United Kingdom in 1986 (see above), and WRK4U in New Zealand in 2003 (see Chapter 4) gave near-immediate estimates of their impact and the decision to expand these programmes could be made after just a few months of highly positive early results. When employment services are outsourced within a consistent framework, outcomes vary substantially, so it is possible to identify the relative impact of different providers fairly accurately without recourse to very complex statistical analysis (Box 5.1). Even in the context of evaluating programmes through random assignment experiments, findings take the form of differences between pilot group (or treatment group) outcomes and control group outcomes, which are not essentially changed by more complex analysis. In general, although the most sophisticated evaluation techniques are usually only applied by academic researchers, the Public Employment Service should be using some form of evaluation as a management tool. At the same time, when the PES uses its evaluations not only for internal management but also to argue for budget allocations, external oversight and verification is needed.

C. The PES should track employment outcomes and not only benefit caseloads

Government often looks to the Public Employment Service to reduce total unemployment because the unemployed are seen as the main target group for its

Box 5.1. **Variability of outcomes under quasi-market arrangements in Australia and the United Kingdom**

Under quasi-market arrangements, the most important employment services delivered to an unemployed person – particularly the case-management function, which includes job-search counselling and assistance – are delivered by competing private providers. Where competition takes place on a “level playing field”, comparisons of client employment outcomes between providers show the impact of more or less effective service provision. Variability of outcomes is greatest in a recently-created market, because by the time the quasi-market has stabilized, relatively poor performers have been eliminated.

In the first tender round of Australia’s Job Network, paid outcome rates (paid outcomes were typically entries to employment lasting at least three months), as measured six months and more after individuals had entered Intensive Assistance services (services for disadvantaged jobseekers), varied within a typical region from about 25% for the highest-performing provider to 9% for the lowest-performing provider (OECD, 2001b, note 80).^{*} The providers from the first contract period that were awarded new contracts in the second tender round had average outcome rates nearly 25% above the overall average of providers in the first contract period (OECD, 2001b, p. 188).

In United Kingdom, the long-term unemployed in selected particularly-disadvantaged urban areas are referred to Employment Zone (EZ) providers, which take over responsibility for employment counselling and placement service from the public provider (Jobcentre Plus and New Deal 25 Plus) for six months. Providers are motivated by a payment system that rewards getting clients off benefits and achieving entries to work that last for at least three months. The quantitative evaluation of this programme (Hales *et al.*, 2003) found that approximately 11 months after each person first became eligible for referral, 34% of EZ participants had experienced a spell of paid work compared to 24% in the control group served by the public system.

As these figures suggest, a high-performing employment services provider may well be able to achieve a 50% increase in employment outcomes for relatively disadvantaged groups: this makes performance management on the basis of measured outcomes feasible without the use of very complex techniques.

Experimental evaluations show similar impacts from some high-performing public programmes. In the United Kingdom, evaluations found that Supportive Caseloading (1993), 1-2-1/Workwise (1994-96, for 15 to 24 year-old long-term unemployed) and 1-2-1 for the Very Long Term Unemployed (1996-97) raised employment rates 26 weeks after random assignment from 8% to 22%, from 12% to 18% and from 8% to 14% respectively. In each experiment, the key additional service was several meetings with an individual case manager (Employment Service Research and Evaluation Branch Reports Nos. 95, 109 and 115, which are briefly summarized in Greenberg and Shroder, 2004).

^{*} Out of 29 regions, 6 had a provider in the one-star category (less than 6% outcome rate) and 15 had a provider in the five-star category (more than 25% outcome rate); performance within a given region usually varied across most of this range (DEWRSB, 2000).

services. By contrast, employment rates get less attention, and in any case the PES is unlikely to be credited with improvements in them.⁹ However, an exclusive focus on reducing unemployment (or reducing caseloads, in the US welfare system) is dysfunctional. Employment services may already be improving the government’s financial balance more through increased tax receipts than through reduced benefit costs. Often, employment services play a role in exposing undeclared work and pushing it back into the declared

economy. However, these outputs are not regularly documented. Long-term tracking of employment and earnings outcomes can bring the additional tax receipts generated by employment services more systematically onto the management radar screen. It can demonstrate that employment services which increase employment and earnings – even if they are typically more expensive than services that only reduce benefit caseloads – generate net benefits for government finances.

In some OECD countries and generally in non-OECD countries, benefit coverage of unemployment is low and a focus on reducing benefit caseloads has little relevance. Objectives such as increasing the “transparency” of the labour market and job-match quality, providing career guidance and vocational training, and promoting formal rather than informal work, are more important. In these circumstances, the impact of employment services cannot be measured in terms of reductions in benefit caseloads, but measurement in terms of impact on employment and earnings outcomes remains relevant.

In most countries, until recently the PES has only been able to track regularly the benefit status of programme participants. Statistical information on *employment and earnings* outcomes achieved by programme participants has been limited to what is known from occasional – often very occasional – questionnaire surveys. However, the technical barriers to data matching are now low, and for research and evaluation purposes OECD countries are increasingly matching the government’s benefit payment databases to its databases of social security contributions and tax records.¹⁰ This raises issues of confidentiality, because the authorities do not want to publish individual earnings data. Access to employment and earnings data by name needs to be highly restricted, while data released for relatively widespread use (for example, by researchers) is made anonymous. Nevertheless, a secure use of employment and earnings records for monitoring the outcomes from labour market programmes at quite limited cost should be possible:¹¹ governments will find it relatively difficult to systematically promote objectives such as employment retention and earnings advancement if it remains difficult or expensive at the operational level to know whether such outcomes are occurring. It is possible to track employment and earnings outcomes without using tax records, but even cumbersome procedures may only generate relatively incomplete information.¹²

D. Defining the value of client outcomes: $B + tW$

Transitions from unemployment to employment are worth approximately $(B + tW)$, where B is (the saving in) benefit payments that results from employment,¹³ t the tax rate on earnings and W earnings, a measure which combines months of employment with earnings per month. $(B + tW)$ can be interpreted as a measure of the impact that increased employment has on the government’s financial balance:¹⁴ the use of programmes whose impacts on $(B + tW)$ exceed their costs improves the government’s net financial balance.¹⁵ $(B + tW)$ is not an exact measure of the benefit from employment services – for example it does not directly incorporate non-wage factors in the quality of job matches or capture how well activation measures achieve sorting of benefit claims (see Chapter 4) – but it seems to be most appropriate measure available operationally.¹⁶ Indicators such as job placements or three-month employment outcomes, which are currently often used,¹⁷ need to be seen as intermediate indicators – accurate only to the extent that programme impacts on them more or less accurately proxy long-term programme impacts on $(B + tW)$. Because the average employment rate for a group of jobseekers who have been referred to

a particular programme typically increases through time, earnings are an increasingly important part of $(B + tW)$ when outcomes are measured over a relatively long period.

E. More institutional conditions

Integrity of benefit and tax administration

When a labour market programme's output is being measured and rewarded in terms of impact on benefit payments and earnings, the integrity of benefit administration needs to be independently monitored and guaranteed. A reduction in benefit costs that results from arbitrary denials of entitlement would not be a useful output: it would be more akin to a form of "gaming" that distorts the outcome measure. Similarly, an increase in declared earnings, as reported in administrative data sources, is not an improvement in outcomes if it results from artificial salary declarations or predatory tax assessment practices.¹⁸

In a quasi-market system, this principle implies that although providers need to be able to initiate benefit sanctions when a jobseeker does not meet conditions (*e.g.* fails to attend an interview, or refuses suitable work), government needs to manage an independent system of tribunals and higher-level appeal courts that protect the rights of jobseekers who appeal.¹⁹

Clear allocation of responsibilities

Positive labour market outcomes may reflect many different types of input. However, it is technically difficult to separately measure the impact from many different types of input.²⁰ Therefore, performance management has to involve allocating responsibility for clients to identifiable units which have relatively long-term responsibility across a substantial range of labour market interventions, rather than splitting responsibility excessively across different programmes, different levels of the hierarchy or different institutions within the PES. When particular units have relatively broad responsibility, they have a large and clear impact on outcomes so that performance management in terms of measures of impact is viable.²¹

This is one reason why PES institutions need to be "integrated". If local employment offices are dependent on practices of a separate local benefit administration in terms of sanctioning jobseekers who fail to attend or if they face erratic variation in the availability of ALMP slots supplied by a third party, only a fraction of local variation in outcomes can be accurately attributed to local employment offices, and then management on the basis of results will not be justified. When employment offices have significant control over the three main functions of the PES, they are able to implement a coherent strategy which has a clearly identifiable impact on these outcomes.²² Nevertheless in a performance management perspective, complete integration of all functions at the same (perhaps local) level is not desirable. Responsibilities also need to be split between local-level "agents" that implement policies while a higher-level "principal" retains responsibility for important areas such as uniform application of benefit eligibility criteria and consistent procedures for the referral of clients to providers and the measurement of provider outcomes.

Deterministic referral processes

The process for referring jobseekers to programmes whose impacts are to be measured needs to be outside the control of jobseekers and service providers (otherwise estimates of impact are easily falsified by selection biases, called "creaming" if the service

provider does the selection).²³ Referral processes that are deterministic in this sense can be implemented by randomly assigning clients to providers, but they are also implemented by traditional arrangements where jobseekers in each local area can only register with one local employment office. Referral processes that are deterministic can allow jobseekers and programme staff to negotiate over the detailed content of services (e.g. the individual's choice of option in the UK New Deal), but not the initial referral (e.g. the individual's choice of New Deal provider). Referral processes that are deterministic can also allow providers or programmes to specialise – for example, a provider might specialise in services for older workers laid off from heavy industries, as long as the providers or programmes then accept all clients from that group who are referred.

As well as preventing spurious variation in outcomes due to “creaming”, performance measures must adjust gross outcomes for exogenous factors that vary across providers for reasons beyond their control (e.g. differences in local client characteristics and labour market conditions). The influence of exogenous factors may be eliminated through random assignment of clients to providers or kept relatively small in other ways (in pilot studies of new programmes, differencing of outcomes between the pilot areas and other areas is often sufficient to control for the influence of exogenous factors). When the influence of exogenous factors is greater, adjustments for it may be made econometrically (e.g. using non-experimental matching methods to evaluate programmes, or using a regression-based “star rating” system to evaluate providers as in Australia).

F. Making evaluation continuous

The welfare system in the United States is now heavily decentralised to states and even decentralized within states, and some observers have noted the limitations of a strategy of governance through occasional impact evaluations and dissemination of their findings. In particular, as Chapter 4 has noted, findings about which type of programme is most effective can vary because important features of programmes are difficult to document and because local contexts vary. From a technical point of view this problem might be solved with more data but this approach may not be feasible, as Greenberg *et al.* (2003) recognize: “Although previous multisite evaluations cannot tell us very much about underlying production functions, future evaluations can. But our analysis suggests that success could take more sites – and more heavy-handed control by the federal government – than is feasible.” A US state might conclude that, once it has learned the basic lessons of a “work-first” approach, national-level evaluations of labour market programmes do not help much with its more detailed problems of governance. Gais (2000) comments that: “Nothing is *finally* implemented; policies may be created and adapted *all the time*... the dynamic character of these systems suggests that, just as there is no final implementation, there may be no final evaluation. Evaluation, to make sense under the new circumstances, ought to be continuous or at least recurrent and built into the management process at the level where critical decisions are being made.”

3. Quasi-market and traditional organisation of employment services

There are several reasons for thinking that quasi-market arrangements can be effective. One is that some local-level administrators know from experience the impact of different measures, and how to use available resources to achieve given objectives. Formal programme evaluation findings do not necessarily provide them with many additional insights.²⁴ A second reason is that quasi-market arrangements implement a “survival of

the fittest” mechanism. Management teams that have good insight into the potential impact of different programmes can respond appropriately even in a public system. But only a “survival of the fittest” mechanism will systematically generalise a successful strategy even when it is difficult to identify which key feature is making it successful. A third factor is that, conditional on successful prior experiences with this approach, quasi-markets could give good results even where government capacity to manage in the complex field of active labour market policy falls short. For example, in US states that now focus welfare policy on short-term caseload reduction, a quasi-market as described here would help ensure that employment services do “more” than caseload reduction.

Despite these arguments in favour of quasi-markets, experience with them in other fields shows that they can go wrong in various ways. Plausibly quasi-market arrangements will be highly effective if the measurement of outcomes and impacts is implemented quite accurately. Australia’s “star ratings” arguably have been measuring provider impacts on sufficiently relevant measures of outcomes sufficiently accurately to achieve reasonable results.²⁵

A. Quasi-market organisation

Full quasi-market arrangements

To implement a quasi-market, the Public Employment Service has to be split between a public authority (the “principal”, here called the government or the purchaser) which determines individual eligibility for benefits and services, assigns clients to a specific provider, and measures outcomes; and multiple employment service providers or local employment offices (the “agents”), which deliver other employment services. The service providers are given near-complete freedom to choose their procedures and programmes, but the purchaser measures the employment outcomes achieved by their clients and in some way ensures that providers are replaced if their outcomes fall systematically below benchmark levels.²⁶

As mentioned above, clients need to be allocated to providers by the purchaser in ways that limit “creaming”. If random assignment methods are used to allocate clients across multiple providers operating in the same local labour market, relative outcomes will measure relative impacts directly or after only minor adjustments. In Australia, clients are allowed to choose a provider but not all of them exert a choice, and those who do not are approximately randomly assigned, which reduces the scope for active “creaming” by providers. It may also be possible to operate a quasi-market with only one provider per locality, using regression adjustments to estimate impacts from gross outcome data.²⁷ However it is not clear whether regression models can measure impacts sufficiently accurately in this case. To minimize problems such as the long-run endogenization of the benchmark,²⁸ some additional procedures – such as contracting for several localities as a package, so that exogenous factors average out, or occasional rotation of providers so as to allow benchmarking by comparison with the preceding provider – can be imagined, but they may remain too costly or inconvenient.

Employment service providers should be able to finance additional services on the basis of their impact on $(B + tW)$. For example if t is 25%, providers should have an incentive to spend USD 1 on additional employment services if this either reduces benefit payments to clients in later years by USD 1, or increases clients’ total earnings in later years by USD 4.²⁹ “Later years” would need to be at least several years, supposed here to be five years.^{30, 31}

One way to implement this arrangement is to actually pay providers the value of client outcomes ($B + tW$). In this case the benchmark would take the form of a fee per client referred, set at a level that allows providers to make only normal profits.³² This fee per client would need to be exogenous to individual provider's employment outcomes (so that incentives for spending on employment services are not distorted), but endogenous with respect to average outcomes of all providers in the longer term (to ensure that employment service providers do not enjoy economic rents, or all make losses). In principle the level of this fee could be determined by a bidding/tendering mechanism conducted separately in each locality: this would remove the need for the government to set the level of the benchmark locality by locality.³³ Then entry and exit from the quasi-market could be based only on provider profitability: more efficient providers would enter the market in a new locality by outbidding the incumbent provider(s) (i.e. offering to handle a batch of new clients with the same formula for outcomes fees, but for a lower level of the fee per client referred). Providers which were less efficient in getting clients off benefits and into work at reasonable cost through their provision of employment services would not be able to win bids at a price that leaves a profit, and would be driven out of the market.

However, paying providers the full value of client outcomes ($B + tW$) over five years – relative to benchmark levels which allow only “normal” profits on average – would subject them to high levels of risk. In the case of small providers risk can imply bankruptcies, which impose additional costs on clients and government. A different arrangement would be that the government pays providers a fixed fee per client to cover the cost of employment services (and normal profits) with no further payments related to outcomes, but tracks the values of ($B + tW$) being achieved by each provider and only renews contracts with the providers that are achieving the best impact. This arrangement eliminates risk (other than non-renewal of the contract) for providers. However, it only generates an optimal level of total spending on employment services if government sets the fixed fee per client at the right level. Also – given that the fee needs to cover the cost of employment services to each client for five years – it allows providers to make profits by providing minimal services for up to five years before being eliminated from the market because of their poor performance. Given these issues, it may be optimal to manage a quasi-market using a mixture of several incentives and safeguards: combining pay-for-results and the principle of selective contract renewal³⁴ with arrangements for more rapid elimination of providers whose performance is exceptionally poor and regulations that enforce minimum levels of service provision.

To the extent that long-term employment outcomes are measured and rewarded, an important practical issue is to implement an “up-front payment principle”. Although the total payments finally received by a provider for a particular set of clients should be ideally based on their unemployment and employment outcomes over a long period following referral (with adjustments only for exogenous factors), advance payments could be made based on all information, within practical limits, that is currently available about the likely final value of these outcomes.³⁵ In this way, achievements such as placements into stable jobs could be rewarded immediately, subject to penalties which claw back the reward if the job later turns out not to be stable. An accurate system of advances would make it easier for providers to invest “now” in programmes which produce employment and earnings outcomes “later”, and would make provider cash flow (excess of outcome payments over operating costs) more useful as a short-term indicator of whether service provision is being successful.

Experience with quasi-market approaches is limited, but the risks include:

- Transaction costs can be high – these are both the costs of contract management for both parties, and costs at the level of individual clients (transfer of clients from the public gateway to the private provider, and continuing interaction between the two in certain circumstances).
- If either poor-quality outcome measures are used or methodologies for determining benchmarks are inadequate, outcomes may be far from optimal.
- Employment service providers may adopt techniques which improve outcomes as measured but not in a substantive sense (“gaming”). However, if benefit and tax records are used as the basis for outcome measurement, “gaming” is unlikely because clients will appeal against unfair reductions in their benefits when they are truly unemployed, and clients will not pay social security contributions or tax if they do not truly have earnings.
- Providers may be able to devise strategies (such as vacancy hoarding) that improve outcomes for their own clients, but impose negative externalities on the clients of other providers.³⁶ The government needs to detect and ban (or perhaps tax) the use of these strategies.
- The public authority (purchaser) may be faced with a “black box”, i.e. it may lack knowledge of what providers are doing. This may limit its ability to identify and control “gaming” behaviour or negative externalities (described above), or make it more difficult to identify and disseminate good practice.
- A quasi-market that rewards the achievement of long-run employment outcomes may tend over time to be dominated by a limited number of fairly large organisations which are able to invest and implement complex strategies, each resembling a traditional PES but operating within a market framework. The market may then become oligopolistic, calling for preferential measures to keep the door open for newcomers.

Despite this long list of potential risks, experience in Australia already shows that all of them are reasonably manageable.

A quasi-market within government?

In principle, quasi-market arrangements could function within government. In this case each local employment office would be run as a (virtual) “profit centre” where income is the value ($B + tW$) of client outcomes (relative to benchmark levels) and outgoings are staff salaries and other employment service costs. Central government would use profits on the (virtual) accounts of these profit centres as its preferred performance measure. However, relatively simple implementations using management-by-results principles for rewarding good performance – e.g. giving performance-related pay to successful employment office managers – may be far different from the operation of a true quasi-market. In Australia and the Netherlands, successful providers can be fairly large organisations. Good performance is generated by successful management structures and business strategies, and efficiency gains arise when responsibility for a particular locality is reallocated from a less-successful organisation to a more-successful one.

Limited subcontracting on the basis of tracking of outcomes

Perhaps a more fruitful use of any system that accounts for outcomes at the level of PES local employment offices is to subcontract employment service provision on an

experimental basis in selected areas, as is done in Employment Zones in the United Kingdom. When the government has information systems that can predict levels of off-benefit and employment outcomes at a particular local office – for example, average values of these outcomes over the subsequent two years, for people who have just entered long-term unemployment – it can invite tenders from private providers to provide employment services for this group on more favourable conditions (i.e. at lower cost if the same employment outcomes are achieved, or for the same cost if better employment outcomes are achieved). If some private providers agree to operate under these contractual conditions, the government can continue tracking client outcomes after clients have exited from the private provider’s services, to check whether the short-term improvements in outcomes they have obtained are sustained in the longer term. As long as a “level playing field” between government and other providers is maintained, this method appears to be a realistic option for the partial or progressive implementation of quasi-markets.

B. Traditional PES organisation and Management by Objectives (MBO)

The Public Employment Service is traditionally a national, hierarchical organisation. This could solve the governance problem using the following principles:

- The PES maintains a national staff ethos. Managers are offered a career with rotation between localities, and potential progression to regional and national management level.
- PES procedures are continuously reviewed and developed through high-quality impact evaluations of existing and potential new programmes. Three main methods are available for evaluating the direct impact of programmes on their participants, each with specific advantages (Box 5.2).
- Best-practice procedures are written into the national “procedures manual”. The national staff ethos and incentives for managers promote compliance with the manual.

Conditional on an ongoing commitment to evaluation and the replacement of programmes that have little impact by more effective ones, traditional PES arrangements have some advantages as compared to quasi-market arrangements. They can partly avoid the institutional constraints and transaction costs that arise from the strict separation between the provider and purchaser roles that is needed to operate a quasi-market. They can potentially implement an approach where multiple types of inputs are evaluated, e.g. strategies for individual employment counsellors, local office characteristics, and specific procedures such as vacancy display or the offer of vocational guidance: in principle the national PES can act rapidly to exploit evaluation findings at any of these levels, even if it is not clear that the average traditional PES acts rapidly in practice. In a quasi-market the impact of each provider (or each local office of each provider) is evaluated as the basis for managing the market, but detailed provider strategies remain mainly inside a “black box”, with a risk that best practices might only spread slowly for that reason.

Public Employment Services in many European countries use “Management by Objectives” (MBO), as described by Mosley *et al.* (2001). Typically, the most important outcome measured is placements (placements into PES job vacancies, as reported by local employment offices without external verification or checks on the duration of the job),³⁷ and outcome levels are compared to targets which are determined by *ad hoc* methods.³⁸ Perhaps related to these weaknesses in the implicit system of impact measurement, MBO systems generally do not prescribe specific action by higher levels of management when

Box 5.2. Three methods for the evaluation of labour market programmes

Research and controversy over the validity of experimental and non-experimental evaluation methods continues. But the three main evaluation methodologies in use each have characteristic strengths.

Random-assignment experiments

Random-assignment experiments appear to often accurately report the impact of services provided to a treatment group, subject to sensible interpretation of the findings in the presence of phenomena such as “control group crossover” (when members of the control group receive the same services as the treatment group). In the case of training and similar programmes in which only a small percentage of jobseekers participate, motivation effects may be small because programme participation is voluntary, or they may be considered unimportant because the focus is on outcomes only for the individuals who participate. In the case of broad strategies which apply to all or most jobseekers much of the interest is in the impact on aggregate outcomes, the fact that random-assignment experiments do not measure motivation effects arising before random assignment, or those which affect the control group, may be important. Good random-assignment practice will attempt to minimise biases (*e.g.* by screening the control group from the expectation of treatment) and occasionally check their size (for example, using techniques similar to AM, 2000, or reworking the random-assignment design to include control sites as well as control groups at a given site).

Non-experimental estimates

Non-experimental impact estimates have many of the limitations of random-assignment experiments, with the additional risk of selection bias and erratic results when complex estimation techniques are applied with no assurance that the underlying assumptions are valid. But they also have important advantages. It is increasingly possible to cheaply estimate impacts for multiple programmes on a continuous basis. Using large longitudinal databases that combine individual information on outcomes, programme participation, and some personal characteristics, national administrations can generate estimates of programme impact without disruption to their regular operations. This can allow estimation of impact for a wide range of programmes and even tracking of changes in the estimated impact of a given programme, in parallel with tracking of its outcomes.

Non-experimental methods can often identify the most successful programmes because their impacts are large. For example, the large impact of Ireland’s Employment Action Plan (Corcoran, 2002) would be hard to miss by any estimation method. Similarly (as noted in Box 5.1) for highly disadvantaged groups of unemployed which are achieving less than 10% employment rates a certain number of months later, it would not be unusual to find that the most successful programmes double this employment rate. Non-experimental estimates with participants and non-participants matched on just a few criteria (*e.g.* age, sex, duration on benefit, and education) can then give approximately correct estimates of impact.

However, selection bias is often an important issue. Non-experimental methods can probably never give a meaningful estimate of the impact of programmes that involve entry to a private-sector workplace. In a situation with no hiring subsidy, hiring is a stochastic event (*i.e.* an event that is not entirely explained by other exogenous or predetermined variables) that has a positive impact on the individual’s subsequent employment history. If we imagine a hiring subsidy that is paid automatically during the first few months of any employment spell that follows unemployment, its “participants” will have a relatively favourable subsequent employment history (after controlling for individual characteristics, etc.) even when the rate of subsidy is zero. The participants in any kind of hiring subsidy or

Box 5.2. Three methods for the evaluation of labour market programmes (cont.)

on-the-job training programme have already covered part of the distance to a regular job – having found a workplace within commuting distance, and identified an employer who expects to be able to work with them. A meaningful estimate of impact can be obtained from experiments where the *offer* of a subsidy is randomised, which has been done occasionally with a finding of modest or even negative impact (Burtless, 1985; Galasso *et al.*, 2002). But this gives an estimate of impact on the population that is given entitlement to the subsidy, not on the individuals that are actually hired with the subsidy. Similarly, in Chapter 4, Chart 4.3, outcomes for language courses are particularly poor. But selection into this programme probably occurs on the basis of factors such as client choice (perhaps some individuals want language training more than a job) or lack of fluency which is observed by employment counsellors but is absent from the researcher’s data set.

Selection biases will tend to bias downwards estimates of impact for programmes that are targeted on barriers to employment. This could cause a systematic tendency for programmes for the disadvantaged to be dropped even when they in reality have as much impact as other programmes. This makes it particularly important from a policy point of view to avoid this type of bias. In the short term, the plausibility of non-experimental estimates needs to be assessed on a judgmental and case-by-case basis (see for example reflections by Jacobson *et al.*, 2004, on the validity of their results). In the longer term, the research agenda needs to include random-assignment experiments or perhaps pilot studies that can characterise the typical size of the selection biases that affect non-experimental estimates.

Non-experimental regression techniques can also model outcomes at the level of local PES offices. Subject to data availability, a regression of PES office outcomes on local office strategies and exogenous economic environment variables generates estimates for the impact of different strategies (information that is used in a hierarchical model of PES management), as well as the additional impact achieved by individual offices for reasons that are not identified (additional information that is used in a quasi-market model of PES management).

Pilot studies

When the government experiments systematically at local office level to identify the impact of programmes – for example, implementing individual action plans after six months of unemployment in some offices but after twelve months in others – it is conducting a pilot study.

In some literature, pilot studies would be described as a particular type of random-assignment experiment (“cluster randomization”). However in an employment policy context, often formal randomization is not necessary. Experimental implementation of a policy change in just a few local offices (chosen to be approximately representative) is often sufficient to estimate impacts. Key outcomes such as the average duration of unemployment spells at one local office relative to the regional average are typically quite stable through time. If outcomes at pilot offices improve soon after the pilot programme is implemented, that can be evidence of impact at a high level of statistical significance. It seems incorrect to suppose or imply that evidence from pilot studies is less accurate or scientific than evidence from individual-level random-assignment studies.

Random assignment experiments often attempt to report the “absolute” impact of a programme, as compared to a control group that receives no services: this may have advantages when, for example, comparing experimental findings across countries. Pilot studies usually take average existing practice as the “control” situation: this will often be more relevant from an operational point of view.

Box 5.2. Three methods for the evaluation of labour market programmes (cont.)

Pilot studies at the level of individual employment offices have some other advantages over classic experiments with random assignment at the level of individuals. They can document the impact of office-wide reforms affecting all jobseekers, e.g. a switch from notice boards to computer terminals for vacancy display. Externalities at local level which affect the control group in a random-assignment experiment (which may be negative, e.g. if increased job-search assistance for the treatment group reduces the number of vacancies available for the control group; or positive, e.g. if the new requirements on the treatment group have a spill-over motivation effect on the control group) are internalized when a treatment is implemented at the level of employment offices as a whole. And pilot studies where a training obligation, for example, is implemented in one locality but not another could measure its total impact including motivation effects, not only impacts on those who are directly referred to it or participate in it.

In pilot implementations, outcomes are sometimes tracked for only a few months because the programme is soon implemented more widely. Also, pilot studies tend to be managed directly by the PES, which may explain why they are rarely used to evaluate existing programmes and sometimes are not written up and published. However they are often feasible and offer good prospects for accurate and relatively cheap estimation of impacts.

measured performance is poor. Local employment office managers still have to follow many PES procedural guidelines, even if they in some cases consider them detrimental to their measured performance. MBO systems are partly effective, but they might be made more effective by clarifying the scope for autonomous decision-making by local management, and moving away from *ad hoc* measures of output and towards a public-sector version of measurement techniques that are robust enough to be used in managing a quasi-market.

Conclusions

In most OECD countries, performance management principles are not applied to labour market programmes in a systematic way. Yet without effective performance management, expensive programmes that have no impact can continue to operate indefinitely. Improvements in labour market outcomes are generally available through more systematic implementation of performance management principles.

Given what is known about programme impacts, OECD countries should, where possible, match benefit data with tax data so as to be able to track long-term employment and earnings outcomes from their programmes at low cost, while assuring individual data protection. As long as benefit reciprocity is the main outcome regularly tracked by the PES, management is liable to focus on achieving off-benefit outcomes rather than long-term employment and earnings outcomes. This is dysfunctional, insofar as the additional costs of programmes that increase earnings can be offset by increases in tax receipts and social security contributions.

Employment and earnings outcomes from employment services can be measured even in developing countries where there is no unemployment benefit system due to widespread informal employment. Performance management of employment services will then, among other things, ensure that employment services are agents promoting the transition from undeclared to declared work.

Notes

1. Nine pilot implementations of Restart began in January 1986. On the basis of weekly monitoring figures, the pilot evidence in February already suggested that a national scheme would cause roughly 23 000 extra people to leave the unemployment register each month. In March, it was announced that the scheme would be implemented nationally as from July. In January 1987, “rolling Restart”, under which the long-term unemployed would be interviewed every six months, was introduced (Price, 2000).
2. As described by Hunn (2000), despite extensive successes the new organisation Work and Income found itself “the object of severe criticism and ridicule around the country... Some of [the criticism] has stemmed from the ‘shoot the messenger’ syndrome: work-first and benefit reductions are not universally popular. Some of it derives from disagreements during both the design and implementation phases which have yet to be settled”. The emphasis on management using Key Performance Indicators (KPIs, in particular, stable employment outcomes) was an issue which “many have raised with the Review Team” and which generated “considerable feeling, amongst staff, purchase and monitoring agencies through to beneficiary advocacy groups... Staff have expressed concern about the strong focus on KPIs in their day to day working lives. There is a view that KPIs do not necessarily reflect the entirety of their workload and that individualising some performance measures makes staff responsible for achieving outcomes outside of their control”. Changes to PES institutions cause disruption and uncertainty, and it is not unusual for performance to deteriorate at first after any major reform. During the second and third tendering rounds of Australia's Job Network in 2000 and 2003, management resources within employment service provider organisations were preoccupied with tender preparation, and the total number of placements achieved by Job Network fell sharply for a number of months (see www.workplace.gov.au – Job Network – Job Network performance statistics).
3. Between 1994 and 2000, Danish labour market policy evolved mainly in the sense that the maximum period of entitlement to benefit on a passive basis was shortened. In 2001, the Danish PES began to implement activation programmes in a more flexible manner in pilot projects in two regions. In 2003, the so-called “75 per cent activation requirement” was abolished in favour of a “stronger focus on an individual approach in employment programmes with a clear job orientation, focus on the shortest way into employment and the involvement of other actors” (see the National Action Plans for 2003 and 2004 at www.bm.dk/english/publications). The new programmes include “interventions in the unemployment spell” as described in OECD (2001a, pp. 41ff).
4. In principle, municipal responsibility for contract management allows experimentation with different methods of contracting (Struyven and Steurs, 2005). However, Australian experience suggests that contract design, evaluation and monitoring is a challenge even for federal governments. Sclar (2000) describes cases where municipalities failed to understand the financial and incentive implications of contractual provisions as well as providers (which are often experienced national organisations), lacked in-house capacity for contract evaluation, or rolled over contracts for many years without effective market testing.
5. Some contracts in the Netherlands now reward outcomes on the basis of “no cure, no pay” (i.e. no fixed fee per client, and payments only for client entries to work). Although these contracts create some incentive for service provision, they are used for groups of less-disadvantaged clients, many of whom will enter work even if no employment services are provided. With this type of contract, a strategy of providing no services can still be profitable for providers in the short term, and long-term survival in the market still needs to be determined by accurately measuring comparative impacts, and not only relying on incentives created by the payment system. As regards outcomes of the Dutch system and what is known about them, a recent newspaper article states that for assistance beneficiaries, the aim for quasi-market employment service providers was to re-integrate at least 40% of those who participated in the trajectories by end 2004. However, of the almost 112 000 “trajectories” that were started (in the largest 30 municipalities) from 2000 to July 2004, just below 20% had led to employment (Trouw, 13 January 2005: www.trouw.nl/nieuwsenachtergronden/artikelen/1105513563971.html). In Australia, placement performance of the Job Network has improved as the system stabilized, so this may happen also in the Dutch system, but the relative lack of direct measures of provider impact may be problematic. Despite low rates of placement municipalities have recently contained growth in social assistance caseloads through anti-fraud and other measures (probably related to the fact that they increasingly bear the full cost of assistance benefits).
6. Cantons with low unemployment rates did not necessarily get a good rating for the performance of their employment offices. Cantons in Switzerland can influence unemployment rates through their offer of places on labour market programmes: in some cantons the offer of places tends to be

made earlier in the unemployment spell and reduce the number of recipients of unemployment insurance while in others places are offered to social assistance beneficiaries and they generate new entitlements to unemployment insurance. Performance ratings of Swiss PES offices may fairly accurately measure the performance of one employment office relative to another, but it is doubtful whether they can measure the average performance of cantonal employment offices separately from the impact of other cantonal policies.

7. Advantages of national-level financing include the mutualisation of financial risk which may otherwise be excessive e.g. for small communities faced with plant closures; ensuring that disadvantaged groups receive support, rather than being banished from the locality; and internalising the benefits of employment services e.g. in the case that worker training leads to geographical mobility. However there is also a case for decentralised financing of benefits in order to ensure that decentralised employment services are cost-conscious. OECD (1994) recommended the retention of a local financing element in social assistance and since then Canada, France, the Netherlands and the United States have transferred social assistance costs, at the margin, to subnational levels of government.
8. Although in Table H of the Statistical Annex in this *Employment Outlook* spending on Categories 2 to 7 (training, job-creation and related programmes) exceeds spending on Category 1 (public employment services and administration), more detailed Eurostat statistics show that about 70% of the spending in the Categories 2 to 7 consists of transfers (e.g. subsistence allowances for training participants and subsidies paid to employers). In terms of services purchased directly, spending on public employment services and administration in European countries is about the same on average as spending on other active programmes.
9. Employment data get less press attention than unemployment data partly because administrative data on unemployment are available with little lag. Employment data, from surveys or administrative sources, when they appear are comparatively old news. And when aggregate employment rates change, it is more difficult to know whether PES interventions are responsible because many of the employed are not former PES clients. So PES impacts on employment outcomes need to be documented at the microeconomic level.
10. Econometric programme evaluations based on data from matched benefit and contribution records are appearing for increasing numbers of countries besides the United States, where evaluations have now used state-level UI contribution records for many years. All the main UK employment and training schemes are now designated for evaluation this way under the Social Security Administration Act 1992 (www.dwp.gov.uk/asd/longitudinal_study/ic_longitudinal_study.asp). In Denmark labour force statistics are largely based on administrative registers (Wismer, 2003), and researchers are able to analyse 15-year and longer records of individual unemployment, employment, training and benefit status (www.grad-inprowe.dk/Economics/kap5-Social.htm). Austria's "Data Warehouse" similarly records employment and earnings, subject to the social security contribution ceiling. In a meeting of experts, European countries without such a system felt there were few technical obstacles: cost could be an issue in some of the countries and "The need to overcome any data protection and privacy issues was considered important, particularly when combining data from different administrative sources, but again was not felt to be a major obstacle if the necessary political will existed". (Peer Review Programme, 2004). In the United States, the National Directory of New Hires, which matches benefit payment and UI contribution records at national level (so that entries to employment in another state are not missed) is now the basis for awarding states the High Performance Bonus for TANF (Wiseman, 2004): this may be the first direct operational use of matched benefit and tax records for performance management (uses for fraud control and evaluation are already fairly common). OECD (2004a) discussed trends in data linking, remarking: "One data match which seems to be lacking or sporadic in most countries is a real-time link between the records of social security contributions (paid on behalf of an employee by the employer) and social security benefits paid to the same person." Since 1990, Australia has computer-matched databases on cycles which must be completed within two months (www.apf.gov.au/library/pubs/bd/1998-99/99bd033.htm), although people do not have unique social security numbers and ongoing matching here could be more difficult. The United Kingdom recently started such a match which led to 80 914 people being caught for benefit fraud last year (*The Guardian*, 8 March 2005).
11. One can imagine a system whereby a regional manager of the PES or the manager of a labour market programme is able to submit a batch of 50 or more social security numbers to a central authority with a statement of why data are needed, and then access key statistics for benefit recipiency rates and total earnings of the batch on a monthly or quarterly basis, subject to statistical safeguards such as random rounding. Techniques such as random rounding can make it difficult to infer individual data from aggregate statistics even by differencing across batches (in

cases where several batches relating to overlapping groups of individuals are issued). Such an arrangement could make systematic tracking of the outcomes achieved by jobseekers who have participated (or are still participating) in particular programmes relatively cheap.

12. In Australia and the United Kingdom, private service providers have to obtain written confirmation from employers to support their claims for payments for initial hires, and again to support claims for three-month employment outcomes. This method of documentation could be extended, but it is already quite costly (some providers employ staff to work full-time on obtaining this documentation), and arguably is not suitable for reporting earnings; and it seems unlikely that traditional PESs will do long-term tracking for performance management purposes by this method.
13. Benefit payments are often liable to some tax or social security contributions but for simplicity these are not mentioned in the formula $(B + tW)$. B can be thought of as the net level of benefit.
14. $(B + tW)$ can also be interpreted as a measure of the net output arising from employment (i.e. the gross output produced, less the disutility of work effort). Grubb (2004) argues as follows: the gain in social welfare when a jobseeker enters work is $W - H$ where W is gross earnings (output) from the job and H is the disutility of hours worked. The jobseeker has an incentive to take such a job if $W(1 - t) - H > B$ where t is the tax rate on earnings and B is the rate of benefit during unemployment. So if jobseekers are involuntarily unemployed (and benefit systems should be managed to ensure that this is the case), $W - H > B + tW$. Therefore the gain in social welfare from an entry to employment is at least equal to $(B + tW)$, which is its net impact on government finances. Note that the condition $W - H > B + tW$ may not typically hold for other types of benefits. Disability benefits, for example, should typically be granted to people for whom the cost (H) of any productive work has become exceptionally high. For this group, although $(W - H)$ may exceed $(B + tW)$ in some cases, this cannot be assumed to be true generally.
15. The $(B + tW)$ criterion is suitable for assessing the value of employment services, but not for example the “making work pay” programmes targeted on low earners discussed in Chapter 3. Reductions in the tax rate t for low earners may be justified for social welfare reasons even when they have a net cost to government (the tax rate t needs to be first set at an optimal level, and then the criterion of impact on $(B + tW)$ can be used to determine real spending on employment services). “Making work pay” measures which involve high marginal effective tax rates for low earners imply a high return to government from employment services that raise earnings in work.
16. The condition for unemployment to be involuntary $W - H > B + tW$ is an inequality and arguably it holds less strongly (and sometimes ceases to hold) at higher levels of B . Also, when a jobseeker finds a job with a higher-than-expected value of W , although this will sometimes be compensation for poor job characteristics (i.e. high H), on average it probably brings $(W - H)$ further ahead of $(B + tW)$. These arguments suggest that job placements should be valued with a weight of somewhat less than 1 on B and a weight of somewhat more than t on W . Additional outcomes that might be rewarded under a quasi-market arrangement include: a) indicators for likely outcomes beyond the end of the direct measurement window (e.g. a bonus if clients upon exit from a five-year follow-up period have acquired professional qualifications or are in a stable and well-paid job); b) indicators of jobseeker disutility and other non-benefit costs arising during the unemployment spell, including penalties on providers for initiating benefit sanctions which turn out to be unjustified or create costs in processing appeals; and c) penalties for nonrespect of regulations, e.g. vacancy hoarding (failure to list job vacancies on the national vacancy database) may be discouraged because it generates negative externalities for the clients of other providers.
17. Contracts with service providers reward entries to employment that last for three months in Australia and the United Kingdom (Employment Zones), six months in the Netherlands, and up to a year (but mainly six months or less) in the United States (contracted-out TANF services)(Grubb, 2004). As long as only such relatively short-term outcomes are being measured, there may be a case for separate financing of investment in certain kinds of training because their long-term impact is thought to be positive, although this case should be checked empirically.
18. If employment service providers are rewarded for earnings outcomes achieved by their clients, they will have a financial interest in increasing the amount of earnings that are declared for the purpose of social security contributions and taxes. Up to a point this incentive is healthy, but there is ample evidence from history and still today (Jenkins and Khadka, 2000) that when tax collection is simply “farmed out” to tax collectors it can become predatory.
19. In the United Kingdom, benefit sanctions are formally decided by Adjudication Officers (who for many years were staff of the social ministry rather than the labour ministry, although these two ministries are currently combined), but on the basis of evidence submitted by an employment counsellor. Getting the appeal system to operate correctly – so that it does provide protection to

claimants, but also does support employment counsellors who correctly impose sanctions – is important for the successful operation of the system of benefit entitlements and activation measures as a whole. One remit for social workers can be to protect and represent groups that may be too weak to defend their benefit entitlements more directly.

20. For example, an individual's positive employment outcome may be attributable to employment counselling received currently, vocational training undertaken three years earlier, legislation which requires particular procedures to be used for job-search monitoring, computer support tools, etc. Some of these inputs are managed at the national level, some at regional level, some at local office manager level and some at individual counsellor level.
21. Another argument for allocating a relatively broad range of responsibilities to employment service providers or programmes is that this allows them to implement “mixed” strategies (combining employment services and longer-term programmes) whose measured impacts are large because they include the motivation effects of referrals to longer-term programmes (see Chapter 4 for discussion of this issue).
22. At the aggregate level (averaging across different local offices), the short-term impact of relatively small adjustments to policy can be evaluated. For example, the United Kingdom has experimentally identified the impact of one additional interview with jobseekers, conducted in the 13th week of unemployment. But in the case of relatively lightweight services (*e.g.* a one-week job-search training course) it must be relatively difficult to accurately distinguish high quality from low quality service providers in terms of long-term impact on employment outcomes.
23. When disadvantaged clients have been definitively allocated to a provider in the sense that the client's employment outcomes will affect the provider's measured performance, “creaming” has been prevented. However, if providers are not sufficiently rewarded or penalized for performance, they may nevertheless choose suboptimal levels of service provision for some clients... a phenomenon described as “parking”.
24. Practical knowledge of “what would work” in welfare reform is described by Mead (2004, p. 197). In Wisconsin, “Policymakers believed they could figure out what worked using pilot programs and their own experience, with little formal evaluation... Once the diversion programmes and more radical work tests came on stream starting in 1994, effects on caseloads and work became much clearer. Administrators could perceive them without benefit [of] research, so evaluations seemed even less justified than before”.
25. Australia's measure of provider outcomes, as used in calculating “star ratings”, is currently based largely on the number of client entries to jobs of at least three months duration with increased weights placed on job entries by the long-term unemployed and very-long-term unemployed (DEWR, 2004). Providers might improve their ratings by delaying their clients' job entries until they have become long-term unemployed or very-long-term unemployed, although monitoring by the employment department shows no evidence of this practice.
26. Benchmark-setting procedures must be as accurate as the procedures that would be used for evaluating the impact (as distinct from the gross outcomes) of labour market programmes: more precisely, actual outcomes minus the benchmark levels must be valid estimates of relative impacts on outcomes.
27. Rubenstein *et al.* (2003) discuss how to use regressions to adjust performance measures.
28. “Endogenization of the benchmark” occurs if a service provider's employment outcomes are regression-adjusted using explanatory variables such as the local unemployment rate, and yet the local unemployment rate is in the long term endogenous with respect to the employment service provider's actions (as it should be, if the employment services are productive). In this case, the method of regression adjustment reduces and potentially eliminates the incentive for providers to actually achieve reductions in the local unemployment rate. It is technically difficult to adequately take into account differences in local labour market conditions without undermining long-term incentives in this way.
29. Note that benefit eligibility conditions in OECD countries allow employment service providers to require participation in public employment services at times when their clients are unemployed (*i.e.* on benefits), but not when they are employed. Payments to employment service providers for increasing their clients' earnings might in principle motivate them to provide employment retention and earnings advancement services during employment. However these payments might also motivate strategies that use the current unemployment spell to deliver earnings-enhancing services, rather than always aiming to shorten its duration.

30. The optimal length of the outcome measurement and provider responsibility period is a matter of judgement. However, the period needs to be more than two years in order to adequately reward providers for delivering various substantive employment services whose employment impact (as documented in Chapter 4) is often zero or negative in the first year, but becomes positive later on. At the same time, if the provider responsibility period is much more than five years, the “market” will begin to resemble one where adults are allocated to a single employment service provider for life. With the period limited to about five years, referrals can be restricted to clients with medium or greater levels of disadvantage (e.g. referral when an individual reaches a threshold of six months of unemployment over the last two years), and clients who enter stable employment can eventually be rotated out of the system.
31. The type of quasi-market arrangements described here would need to be introduced progressively. The purchaser at first necessarily uses relatively short-term measures of employment outcomes, and limits risk for providers who cannot know in detail what market conditions they will face.
32. $(B + tW)$ measured over a five-year period is a very large amount of money and it might be imagined that clients could “blackmail” providers, by saying “please pay me a back-to-work bonus or hiring subsidy, or I will remain unemployed for a long time and cost you a lot of money” (or potential employers might do something similar). But successful providers will establish a reputation for never paying hiring subsidies, except in cases where this is genuinely necessary to achieve an employment outcome and by methods that do not lead clients or employers more generally to expect the same subsidy. Such an approach may be most effectively implemented by skilled employment counsellors operating with considerable discretion.
33. Under ideal conditions (absence of collusive bidding and costless availability of relevant information), auction mechanisms are able to set the correct price for items that have a unique set of characteristics.
34. The government may pay providers a fixed fee per client and renew contracts with the providers which achieve the highest values of $(B + tW)$ relative to benchmark values, but also pay providers a fraction v (e.g. half) of the value of $(B + tW)$, relative to benchmark values. This is approximately what Australia now does (although Australia does not use $(B + tW)$ as its measure of outcomes, and performance is measured in terms of gross outcomes relative to benchmarks for purposes of contract renewal, but in terms of unadjusted gross outcomes for purposes of fee payment). This arrangement guarantees that any service that increases $(B + tW)$ by USD 1 at a cost of USD v or less will be provided, even if the level of the fixed fee per client set by government is suboptimal.
35. For example providers could be allowed to report to government their placements of clients into stable jobs to government and receive an up-front payment corresponding to several months of future employment outcomes (the average number of month that clients stay in these jobs). In the same way, providers which put jobseekers into a training programme that has a record of achieving long-term employment outcomes could qualify for financial advances.
36. At the same time, actions by service providers that reduce unemployment rates among their own clients have positive externalities in terms of motivation effects on individuals who are not or do not become their clients (see Chapter 4 and official evaluations of Intensive Assistance in Australia). So a system that “pays” providers for their impact on outcomes of their own clients may under-reward their actions. Potential positive and negative externalities both need to be kept in mind.
37. Recent “scandals” over placement claims in some European countries (noted by Grubb, 2004) appear to reflect the fact that local employment office staff had incentives to report placements, but with no system of external verification. A simple procedural improvement would be to use New Zealand’s system where, when an employment counsellor claims a placement, this is only validated after the client has been off benefits for at least three months.
38. Commonly in MBO systems the target (i.e. benchmark) for outcomes next year is set as a slight mark-up on the outcomes in the current year. This endogenizes the benchmark from the point of view of any local office manager who stays in place for more than year, undermining the incentive to improve performance.

Bibliography

- AM (Danish Ministry of Labour) (2000), "Effects of Danish Employability Enhancement Programmes", Copenhagen (www.bm.dk/english/-documents-orderpublications).
- Burtless, G. (1985), "Are Targeted Wage Subsidies Harmful? Evidence from a Wage Voucher Experiment", *Industrial and Labour Relations Review*, Vol. 39, October, pp. 105-109.
- Corcoran, T. (2002), "Retrospective Analysis of Referral under the Employment Action Plan (EAP)", FÁS, Ireland (cited at: www.fas.ie/FAS_Review/SF.html).
- Council of Economic Advisers (1997), *Explaining the Decline in Welfare Receipt, 1993-1996*, Washington DC (<http://clinton4.nara.gov/WH/EOP/CEA/Welfare/index.html>).
- DEWR (2003), "Intensive Assistance and Job Search Training: A Net Impact Study", Evaluation and Program Performance Branch Report 2/2003, Australia (www.workplace.gov.au – publications – employment – evaluation of programmes and services).
- DEWR (2004), "Job Network Star Ratings July 2004: Comparative ratings of Job Network members contracted to deliver Job Network services from 1 July 2003", Australia (www.workplace.gov.au – Job Network – performance statistics).
- DEWRSB (2000), "Job Network Member Performance Information as at 31 October 1999", Australia (www.workplace.gov.au – Job Network – performance statistics – performance and evaluation reports archive).
- Dixon, S. (2002), "Using Administrative Data Sources in Labour Market Research: an introduction", *Labour Market Bulletin*, Vol. 2000-02 Special Issue, New Zealand (www.dol.govt.nz – publications – research papers).
- Gais, T. (2000), "Concluding Comments: Welfare Reform and Governance", in Weissert, C.(ed.), *Learning from Midwestern Leaders*, Rockefeller Institute Press, New York, pp. 1-24 (www.rockinst.org/publications/federalism/learn_leaders_chap_7.pdf).
- Galasso, E., M. Ravallion and A. Salvia1 (2002), "Assisting the Transition from Workfare to Work: A Randomized Experiment", World Bank Development Research Group Poverty Team, Policy Research Working Paper 2738, Washington (http://econ.worldbank.org/files/3183_wps2738.pdf).
- Greenberg, D., R. Meyer, C. Michalopoulos and M. Wiseman (2003), "Explaining Variation in the Effects of Welfare-to-Work Programs", *Evaluation Review*, Vol. 27, No. 4, pp. 359-394.
- Greenberg, M. and M. Shroder (2004), *The Digest of Social Experiments: Third Edition*, Urban Institute Press, Washington.
- Grubb, D. (2004), "Principles for the Performance Management of Public Employment Services", *Public Finance and Management*, Vol. 4, No. 3, pp. 352-398.
- Hales, J., R. Taylor, W. Mandy and M. Miller (2003), "Evaluation of Employment Zones: Report on a Cohort Survey of Long-Term Unemployed People in the Zones and a Matched Set of Comparison Areas", DWP/JAD Research and Analysis Publications No. 176, Department for Work and Pensions, United Kingdom (www.dwp.gov.uk/jad/index_intro.asp).
- Hunn, D. (2000), *Report of the Ministerial Review into the Department of Work and Income*, report released by Hon. Trevor Mallard, Minister of State Services, New Zealand (www.executive.govt.nz/minister/mallard/winz/index.html).
- Jacobson, L., R. Lalonde and D. Sullivan (2004), "Estimating the Returns to Community College Schooling for Displaced Workers", IZA Discussion Paper No. 1017, Germany.
- Jenkins, S. and R. Khadka (2000), "Modernization of Tax Administration in Low-Income Countries: the Case of Nepal", CAER II Discussion Paper, No. 68, Harvard University, Cambridge, MA (www.cid.harvard.edu/caer2/htm/).

- Johri, R., M. de Boer, H. Pusch, S. Ramasamy and K. Wong (2004), *Evidence to Date on the Working and Effectiveness of ALMPs in New Zealand*, Department of Labour, New Zealand (www.dol.govt.nz/browse-dol.asp).
- Maré, D. (2002), "The Impact of Employment Policy Interventions", *Labour Market Bulletin*, Vol. 2000-02 Special Issue, New Zealand (www.dol.govt.nz-publications-research-papers).
- Mead, L. (2004), *Government Matters: Welfare Reform in Wisconsin*, Princeton University Press, Princeton.
- Mosley, H., H. Shütz and N. Breyer (2001), *Management by Objectives in European Public Employment Services*, WZB Discussion Paper FS I 01-203, Social Science Research Center, Berlin (www.wz-berlin.de/ars/abstracts/i01-203.en.htm).
- OECD (1994), *The OECD Jobs Study*, Paris.
- OECD (2000), *Employment Outlook*, Paris.
- OECD (2001a), *Labour Market Policies and the Public Employment Service: Proceedings of the Prague Conference*, Paris.
- OECD (2001b), *Innovations in Labour Market Policy: the Australian Way*, Paris.
- OECD (2002), *Employment Outlook*, Paris.
- OECD (2003), *Employment Outlook*, Paris.
- OECD (2004a), *Employment Outlook*, Paris.
- OECD (2004b), *Economic Survey of Switzerland*, Paris.
- Peer Review Programme (2004), "Data Warehouse Monitoring in the Public Employment Service in Austria", Vienna (www.peerreview-employment.org/en/austria04/AUS04.htm).
- Price, D. (2000), *Office of Hope: A History of the Employment Service*, Policy Studies Institute, London.
- Rubenstein, R., A. Schwartz and L. Stiefel (2003), "Better than Raw: A Guide to Measuring Organizational Performance with Adjusted Performance Measures", *Public Administration Review*, Vol. 63, No. 5, September/October, pp. 607-615.
- Sclar, E. (2000), *You Don't Always Get What You Pay For: the Economics of Privatization*, Cornell University Press, Ithaca and London.
- Struyven, L. and G. Steurs (2005), "Design and Redesign of a Quasi-market for the Reintegration of Jobseekers: Empirical Evidence from Australia and the Netherlands", *Journal of European Social Policy*, Vol. 15, No. 3 (see also www.hiva.be/docs/paper/P19_LS_Quasi-market.pdf).
- Wallis, J. (2001), "Different Perspectives on Leadership in the New Zealand Public Sector: the Curious Case of Christine Rankin", School of Business Paper No. 0113, University of Otago, New Zealand.
- Wiseman, M. (2004), "The High Performance Bonus", Paper prepared for the 26th Annual Research Conference Association for Public Policy Analysis and Management, Atlanta, Georgia, October 28-30 (<http://home.gwu.edu/~wisemanm>).
- Wismer, K. (2003), "Use of Registers in Social Statistics in Denmark", Paper for United Nations Statistics Division Expert Group Meeting on Setting the Scope of Social Statistics, New York, 6-9 May (http://unstats.un.org/unsd/demographic/meetings/egm/Socialstat_0503/doclist.htm).
- Zellman, G., J. Klerman, E. Reardon, D. Farley, N. Humphrey, T. Chun and P. Steinberg, (1999), *Welfare Reform in California: State and County Implementation of CalWORKs in the First Year*, RAND, Santa Monica (www.rand.org/labor/CalWORKs/publications.html).