

West London Working

West London Working Evaluation

Employability Performance Rating



Unless alternative contractual arrangements have been made, this report remains the property of Roger Tym & Partners until fees have been paid in full.

Copyright and other intellectual property laws protect this information. Reproduction or retransmission of the information, in whole or in part, in any manner, without the prior written consent of the copyright holder, is a violation of copyright law.

The above table is to be completed in Final Reports only to show the progress for ISO purposes

CONTENTS

	EXECUTIVE SUMMARY.....	i-ii
1	EVALUATION APPROACH	1
	Introduction.....	1
	The methodology we used	1
2	WHAT, WHERE, WHO AND WHICH: THE SERVICE OFFER.....	5
	Introduction.....	5
	What was the nature of activities?.....	5
	What was the rationale for the project?.....	6
	What were the objectives for this work package?	7
	What were the target outputs and outcomes for this work package?.....	8
	Where was the work package delivered?.....	8
	What were the work package timescales?	9
	Which organisations have been involved in delivering the work package?.....	10
	How much has been spent on the work package?.....	11
	What has happened since the completion of the Pilot? How is the Performance Rating is being used now?	11
3	DEFINING HOW THE WORK PACKAGE AIMED TO ACHIEVE ITS OBJECTIVES	13
	Introduction.....	13
	The Vision statement.....	13
	What was the work package logic model (“theory of change”)?.....	13
4	HAS THE WORK PACKAGE DELIVERED WHAT WAS HOPED?	17
	Introduction.....	17
	Has this work package achieved what it set out to?.....	17
	Was there any duplication of provision?.....	18
	Does the Performance Rating system remain relevant?	21
	How has the work package expenditure performed against targets?	22
	Were changes made from the initial concept? Why?	22
	What could have been improved during the development process?.....	23
	Were there any unintended impacts or indirect effects resulting from the work package’s activities?.....	24
	What has worked well and not so well?	25

	How (if at all) does the work package link to other WLW activities and objectives?	27
5	EVALUATING THE STRATEGIC ADDED VALUE ACHIEVED	29
	Introduction.....	29
	What is Strategic Added Value (SAV)?	29
	How we evaluated SAV	29
	Findings on SAV.....	29
6	EVALUATING VALUE FOR MONEY	32
	Introduction.....	32
	Did the work package represent value for money?	32
7	SUMMARISING THE LESSONS LEARNT AND DRAWING CONCLUSIONS.....	34
	Introduction.....	34
	Summary of findings.....	34
	Lessons that inform future approaches	36

EXECUTIVE SUMMARY

1. **West London Working (WLW) is one of 15 City Strategy Pathfinders** launched by the Department for Work and Pensions (DWP) in April 2007. The Employability Performance Rating work package was one of thirteen work packages undertaken by WLW.
2. **The objective of the Employability Performance Rating was to make the “marketplace” for providers of employment services function more efficiently**, so delivering higher quality services. The employability service providers are awarded four stars; the worst receive no stars. Ratings will be made public, so leveraging up project quality. The system
 - Provides information upon which good quality commissioning decisions can be made, both by public funders, and contractors who are looking to sub-contract to smaller organisations.
 - Enables providers to benchmark themselves against each other which would lever up quality through providers’ own efforts to improve their ratings;
 - Supports prime contractors and funders to transparently manage their supply chains; and
 - Allows residents to select high quality provision from the alternatives on offer.
3. The Performance Rating was developed using **best practice experience from Australia, and built on experience from Skills Funding Agency and DWP**. The Performance Rating was developed by West London Working and external consultants through close consultation with stakeholders and users.
4. **The Performance Rating is now being used across London**. The London Development Agency implemented use of the Performance Rating from April 2010 with 12 of its major employment and skills prime contracts, and 32 subcontractors. It is rolling out its use with all of its learning and skills grants from April 2011 which constitutes £20m of annual investment. London Councils implemented with all of their ESF funded provision from October 2010, incorporating a further 30 providers across the capital. The Greater London Authority will sponsor the use of the Performance Rating in future years.
5. **The logic model’s “theory of change” remains effective as a piece of logic**. We see the strategic rationale for the development of the Performance Rating system perfectly: that, in a very crowded “marketplace” of providers of employability services in West London, there needed to be a way of funding the best providers and avoiding funding the worst, and allowing providers to benchmark quality. As such, the Performance Rating system was very sensible response to a classic “information” market failure.
6. **However, the way that the logic model’s “theory of change” plays out in the real world has yet to be seen**. The evaluation found that the system was well conceived and well delivered through the development process, but real evidence that the Performance Rating system was working would need to be seen across the network of provision as a whole. In many respects the Performance Rating system has not yet been able to prove its worth in a system-wide way, although, having interviewed stakeholders and tested the logic underlying the project, the evaluation found that that its success would be likely.
7. **However, there are questions about whether the Performance Rating system will be given the chance to prove itself**. The problem is that circumstances have changed. The

Performance Rating system is in the process of losing its sponsoring organisation (the LDA). Whilst the GLA will sponsor the Performance Rating system in the absence of the LDA, the GLA has not made the same organisational investment in the system. Most significantly, the Work Programme represents a very significant change in the marketplace for employability services. It will be challenging to get Prime Contractors to adopt the system in the Work Programme without DWP making use a contractual requirement.

8. **It is not possible to make quantitative calculations of value for money of the Performance Rating system. We therefore looked at value for money more qualitatively.** £229k was spent on the project, which was subject to open tender. If the Performance Rating system is shelved, then clearly this will prove to be very poor value for money. If the Performance Rating system is widely adopted, and proves valuable in leveraging up the quality of provision and the quality of management decisions, then the scheme could prove to be very good value.
9. **The positive reaction found by the evaluation suggests that future deliverers of employment services, both inside and outside of DWP's Work Programme, should give serious consideration to using the Performance Rating system.**
10. **Lessons learned** through the evaluation were that
 - The development process worked well and was an excellent example of collaborative development.
 - Future users need to be aware of the debate between the merits of relative or absolute measurements of provider success, and how ratings take account of late issue of grants.
 - To maximise the impact of the Performance Rating system it requires a regional sponsor to co-ordinate take up of the rating and to publish results.
 - Ongoing publicity about how the system can be used is helpful, and self assessment processes and user training could be bolstered further.
 - The evaluation found that future users of the may wish to be more aggressive in their use of the Performance Rating system, as in the Australian model, which integrates commissioning with the Performance Rating.

1 EVALUATION APPROACH

Introduction

1.1 This section describes the method we used.

The methodology we used

We compared different accounts of delivery in a process of “triangulation”

- 1.2 In order to get a balanced and “three dimensional” view of the work package, we used a method which allows us to triangulate the views of the different actors involved at different stages of the process. We also interviewed staff from other City Strategy Pathfinders (CSPs) to further bolster this approach. The process of comparing the accounts provided by different sources allowed us to weigh up different accounts of the process, and get perspectives from different angles on what worked, what did not, and what could have been improved.
- 1.3 We first reviewed the documentation available through the West London Working website. This included working papers, business plans, and staff presentations. Drawing all the evidence together allowed us to reach some initial findings and conclusions as to the overall impact of the Employability Performance Rating.
- 1.4 We then tested and developed our findings further via interviews with key people. We were careful to include a number of different organisational perspectives here, in order that different views could be compared. Altogether we undertook 16 interviews, undertaken via a mix of face-to-face and telephone interviews. Eight of these were in-depth interviews.
- WLW
 - LDA (Senior Management)
 - LDA (Contract Managers)
 - DWP
 - KPMG
 - Young People Learning Agency (in view of responsibilities previously undertaken through the LSC)
 - London Borough of Brent
 - London Borough of Ealing
 - London Borough of Hillingdon
 - London European Programme Management Unit
 - West London Network
 - London Councils
 - Ingeus (a prime contractor)
 - Burleigh College (an SME provider)
 - New Challenge (an SME provider)
 - P3 (an SME provider)

We quantified the more qualitative responses

- 1.5 A number of targets are qualitative in nature. It is helpful if we were to convert some of these qualitative notions into a more quantitative scale. We used Likert Scales to capture levels of agreement with specific statements scored on the following scale: strongly disagree (5)/ disagree(4)/ neither (3)/ agree (2)/ strongly agree (1).
- 1.6 These responses are not statistically significant in any way, and as a result we cannot rely too heavily on these findings. However, we can collate different partners' responses to the Likert statements, and use the responses to quantify how successfully the work package has delivered some of these more qualitative goals.

We used a "theory of change" approach

- 1.7 Alongside this process of "triangulation", we used a "theory of change" approach. Simply put, a "theory of change" methodology requires us to understand the mechanism by which the programme is supposed to effect change, and then evaluated whether this mechanism did in fact work.
- 1.8 We see two steps for a theory of change approach to this evaluation. The first step is undertaken in this report, and the second is undertaken in the meta-report.
 - The first step, undertaken in this work package evaluation report, used a theory of change approach to understand why WLW thought that this work package would be valuable, and how the work package would achieve its goals. We then evaluated to see if that mechanism worked by talking to work package managers, delivery staff, other City Strategy Pathfinder stakeholders and partners, and other stakeholders. This provided a model which we used to judge the success of the work package within its own terms. This first stage allows for a close look at the internal logic of the work package and if it has successfully delivered its own objectives, but it does not address how the work package fits with the wider context of the WLW Programme. We therefore need a second step.
 - The second step, undertaken in the meta-report (published separately), explores whether the deliverables of the individual work packages are relevant to the overall objectives of the WLW programme. This stage asks why WLW thought that each of the work packages would deliver its overarching objectives for WLW. This places the individual work packages in a wider context.

We evaluated Strategic Added Value

- 1.9 Strategic Added Value (SAV) is a category of outcome which relates specifically to how strategic programmes use their influencing and lobbying roles to generate positive economic outputs. This recognises that the contribution of programmes can come through their work package and programme spend but also through their influence on partners and stakeholders' behaviour and performance.
- 1.10 Three methods were used here.
 - Firstly, our document review was used to investigate where SAV was sought.

- Secondly, carefully structured interviews with stakeholders were undertaken to assess how far the worksteam has delivered SAV.
- Thirdly, we quantified these qualitative impressions on SAV using Likert statements.

2 WHAT, WHERE, WHO AND WHICH: THE SERVICE OFFER

Introduction

- 2.1 This section provides the basic information on the work package evaluated. In this section we also define what the work package aimed to achieve. Evaluators can only tell if a work package has been successful if they know what it set out to deliver in the first place.
- 2.2 This section simply seeks to provide information. The information provided here is subject to analysis in later chapters.

What was the nature of activities?

- 2.3 This WLW Work Package created the “Performance Rating” system.
- 2.4 A full, detailed explanation of the workings of the system can be found on the West London Working website (see the footnote for the address).¹ In summary, the Performance Rating benchmarks the performance of contractors and sub-contractors engaged in delivering a wide range of skills and employment services in West London. The best performers are awarded four stars; the worst receive no stars. It was the intention that all ratings would be made public, so leveraging up project quality.
- 2.5 The Performance Rating seeks to
- Transparently assess and benchmark the performance of skills and employment providers by using management and performance information collected by delivery partners and funders.
 - Provide an evidence based track record of delivery against grant.
 - Provide a framework for Lead Delivery Partners to performance manage their sub-delivery partners.
 - Support customer choice and collect evidence of customer satisfaction.
- 2.6 Stars are awarded on the basis of a weighted assessment of performance in three key performance areas (KPAs): Contract Performance, Quality and Contract Compliance.
- 2.7 For each KPA, performance is assessed by the use of performance indicators. The indicators for each KPA are set out in the table below.

¹ <http://www.westlondonworking.org.uk/index.php?miid=1668>

Table 2.1 Performance Rating Key Performance Areas

Key Performance Area (KPA)	Weighting of the KPA	Indicators	Weighting of indicator
Contract Performance	60%	• Delivery against contract targets	48%
		• Delivery against contract diversity targets	12%
Quality	30%	• Conversion Factor	12%
		• Self assessment of quality	9%
		• Client Satisfaction	9%
Contract Compliance	10%	• Contract compliance and contractor pro-activity	10%

Source: Guidance for West London Employability Performance Rating v3

- 2.8 Irrespective of the points achieved, four stars (the highest rating) will not be awarded unless the provider delivered 100% on contract targets (i.e. achieved 60 points for the contract targets KPI).
- 2.9 The Performance Rating applies to each contract a provider supplies. The majority of LDA and Local Authority funded contracts are delivered by multiple delivery partners. Performance ratings therefore cascade down through contracts, with the funder rating the prime contractor, and the prime contractor rating any sub-contractors (sub-contractors being rated have to have an annual minimal contract value of £50,000 to be included).
- 2.10 Weightings attached to the indicators were developed in conjunction with the pilot work package Project Board and Development Group, which was made up of contract managers from funders and deliverers of employment and skills services.
- 2.11 The Performance Rating went further than the collation of the usual management information on contract compliance and target delivery. One of the more innovative features of the Performance Rating was that it measured, firstly, client satisfaction (which measured the satisfaction of service users with the provision they received). This received a weighting of 9%. Secondly, the process included a self assessment of quality (which is undertaken by Service Providers rating their own performance). This received a weighting of 9%.
- 2.12 The Employability Performance Rating timetable incorporated a six month review of draft self assessment scores where Contract Managers and or prime contractors were able to request that the Provider reviews its self assessment ratings.

What was the rationale for the project?

- 2.13 The concept of a West London Performance Rating was developed to overcome some of the problems identified in the West London Working Business Plan. Essentially, with 200 providers of employment and skills services in West London alone, provision had become too complex to comprehend, and tools were thought necessary to allow some kind of improved management control.
- 2.14 The WLW business plan suggested that there was

- No central quality and performance management system, which meant that, other than word-of-mouth, there was little way of knowing which contractors should be avoided and which rewarded when awarding contracts, with the resulting churn of poor quality providers in the system;
- Lack of coordination in the provision and planning of services, which meant that contractors and sub-contractors frequently did not know of each other, and had no idea of how well each were performing – with the result that partnering arrangements and other forms of co-operation were obstructed; and
- Multiple and confused points of contact for beneficiaries and employers, which reduced take-up of services, and increased costs.

2.15 The Performance Rating system was intended to be a fix for the first two of these problems, and to assist in the third problem by allowing personal advisors to direct clients to high quality provision straightforwardly. The provision of a self-assessment element was also intended to allow some of the smaller providers to adopt the more sophisticated management techniques of the larger providers.

2.16 The longer term impacts of the system were to create a marketplace for employability services which was more responsive to client and funder demand for high quality provision.

2.17 WLW documentation detailed the key challenges for the Performance Rating with more precision. These help explain the project rationale. The challenges were seen as being²:

- Employability provision has a variety of grant targets in respect of outputs and outcomes;
- There are differences between the level of performance data collected for each LDA Grant Agreement, and that collected for Local Authority (LA) Grants;
- There are different LA and LDA audit and monitoring processes across grant streams;
- Currently, there is no standard process for assessing quality and providers are inspected infrequently by Ofsted;
- The grants have a variety of start and end dates;
- Many of the grants involve sub-delivery partners; and
- Many delivery partners have multiple grants with different funders

What were the objectives for this work package?

2.18 The objective of the Performance Rating system was to make the “marketplace” for providers of employment services function more efficiently, so delivering higher quality services more efficiently. This would be achieved by

- providing information upon which good quality commissioning decisions could be made, both by public funders, and contractors who were looking to sub-contract to smaller organisations.

² WLW (May 2010) *Guidance for the Employability Performance Rating Version 3*

- Enabling providers to benchmark themselves against each other which would lever up quality through providers' own efforts to improve their ratings;
- allowing beneficiaries to be directed towards high quality providers by personal advisors; and
- to allow beneficiaries to select high quality provision from the alternatives on offer.

2.19 There was an ambition to expand the system to cover provision funded by other sources and to enable use by a range of providers in a supply chain such as Local Authority, charity and Regional Development Agency funded providers of worklessness services.

2.20 The development of the Performance Rating methodology sought to use and build on:

- Skills Funding Agency's Framework for Excellence and DWP's Star Rating;
- Existing performance indicators and information used to report the performance of non-Skills Funding Agency and non-DWP funded provision; and
- Management information that is currently collected by Providers and Funders in respect of non-Skills Funding Agency and non-DWP funded provision.

What were the target outputs and outcomes for this work package?

2.21 Outputs are the final quantifiable products produced by the work package for society. Outcomes are the broader impacts or consequences for the community of the activity of the work package.

2.22 There were no specific quantified output targets for this work package. It did not aim to directly assist people into work, or provide training. Instead, the desired outcomes were to create the Performance Rating system, so that it could lever up the quality of employability provision in West London, so contributing to the delivery of the overall objectives of WLW. We have discussed these overall objectives in above.

Where was the work package delivered?

2.23 The work programme was delivered over different geographical areas at different times.

- The pilot phase (summer 2008 to summer 2009) saw the initial development and testing of a "non-live" Performance Rating system with a pilot group. Each of these pilot group members was located in the West London Working area (Brent, Ealing, Hammersmith and Fulham, Harrow, Hillingdon, and Hounslow)³.
- The 'trailblazing' first live pilot version of the Performance Rating system was run during 2009/10 in Hillingdon with the Gateway Heathrow 2012 LDA ESF Grant project.
- Delivery boundaries shifted again during the broader launch phase of the Performance Rating system. The Performance Rating system is intended to go London-wide.

³ The pilot Development Group included Burleigh College, Catalyst Communities Housing Association, Hammersmith and Fulham Council, Hammersmith and Fulham Volunteer Centre, Heathrow City Partnership, Imperial College NHS Trust, LB Ealing Learning Curve, Refugees into Jobs, Notting Hill Housing Trust, Brent Into Work, P3, and Nord Anglia Lifetime Development.

- 2.24 The Performance Rating was developed with the intention that it would be appropriate for use across the UK.

What were the work package timescales?

- 2.25 The procurement process for support to develop the performance rating took place between March 2008 and May 2008. KPMG were appointed. The development process started in earnest in May 2008 with an event which included a wide range of partners including the LDA, SFA, LAs, KPMG, providers and contract managers. The event was facilitated by KPMG, and saw the Key Performance Areas (KPA) set. These KPAs stayed with the project throughout.
- 2.26 Between May 2008 and September 2009, a prototype performance rating was piloted and further developed with the prime grant holders and sub-delivery partners from 11 LDA and one London Borough funded contracts delivering in West London. During that period (between January-March 2009) the Key Performance Areas, and underneath these the key performance indicators, were finalised.
- 2.27 There was a phased implementation of the Performance Rating system during 2009/2010, beginning in November 2009.
- 2.28 A final feedback session to all Pilot Development Group members was carried out in Sept 2009. The final intervention by KPMG was Feb 19th 2010.

Table 2.2 Implementation timetable

Apr 09	LB Hillingdon commits to trail blaze performance rating of Heathrow Gateway 2012 LDA ESF grant.
Oct 09	Request inserted in Flexible New Deal 2 specification for West London prime contractor to utilise the Performance Rating to performance manage its subcontractors
Jan 10	Formal communication to sample 'Phase 2' grant holders that will be requested to use the Performance Rating in 10/11
Feb 10	Performance Rating workshops with funders, prime contractors and subcontractors.
Feb 10	Phase 2 grant holders required to implement requirements of the Performance Rating
Apr 10	Phase 2 grant holders formally commence Performance Rating
Apr 10	All LDA employment focused grant agreements to include requirement to adhere to Performance Rating methodology.
May 10	Results of Heathrow Gateway 2012 prime contractor and subcontractor ratings reported
Oct 10	London Councils formally commence performance rating of London ESF grants
Dec 10	London Development Agency, London Skills and Employment Board and Greater London Authority request that DWP work programme providers use the performance rating and annually publish their supply chain ratings

Dec 10	LDA publishes six month stock take of phase 2 grant holders that commenced using the Performance Rating in April 10.
Feb 11	All LDA learning and skills grant holders required to implement requirements of the performance rating
Feb 11	Performance Rating Workshops held for DWP Employment Related Support Service Framework providers in the London Lot (including potential Work Programme Prime Contractors).
Apr 11	Grants in the performance rating cohort expanded
Apr 11	All LDA learning and skills grant holders formally commence performance rating
May 11	Phase 2 performance ratings collated [LDA and London Councils]
May 11	Phase 2 performance ratings published
May 12	11/12 performance ratings published
May 12	11/12 performance ratings published

Which organisations have been involved in delivering the work package?

- 2.29 The pilot and development of the performance rating was overseen by the West London Working Performance Rating Work Package Board, consisting of representatives from WLW, London Development Agency (LDA), Learning and Skills Council, DWP and the London Borough of Brent.
- 2.30 KPMG was appointed by the London Development Agency (LDA), on behalf of West London Working (WLW), to develop, pilot and implement the performance based rating.
- 2.31 Given the broad spread of organisational linkages required if the project was to achieve its objective of widespread take-up, it was the case that broad buy-in to the project was required.
- 2.32 A number of delivery partners and contract types were selected for the pilot phase from July 2008 to July 2009. This phase included Burleigh College, Catalyst Communities Housing Association, Hammersmith and Fulham Council, Hammersmith and Fulham Volunteer Centre, Heathrow City Partnership, Imperial College NHS Trust, LB Ealing Learning Curve, Refugees into Jobs, Notting Hill Housing Trust, Brent Into Work, P3, and Nord Anglia Lifetime Development.
- 2.33 The role of the pilot delivery partners was to help develop and test the prototype Performance Rating methodology, including
- the definition of each of the proposed performance indicators and their measurement;
 - the definition of assessment criteria/benchmarks against which performance would be assessed; and
 - the additional administrative commitment, for example new data-collection processes.

How much has been spent on the work package?

- 2.34 In total, £226,825 including VAT has been spent on the project. This has been spent on KPMG's fees.
- 2.35 It should be noted that this project has also absorbed considerable amounts of time of WLW staff and other stakeholders who worked collaboratively with KPMG during intensive parts of the project.

What has happened since the completion of the Pilot? How is the Performance Rating is being used now?

- 2.36 The Performance Rating system will be used in the following ways:
- The London Development Agency implemented use of the Performance Rating from April 2010 with 12 of its major employment and skills prime contracts, and 32 subcontractors. The LDA is rolling out Performance Rating to all learning and skills grants in 11/12. Grants in contract from 12/13 will be folded into the GLA structure.
 - The London Skills and Employment Board and Greater London Authority have requested that preferred suppliers on the Department for Work and Pensions' Employment Related Services Framework (ERSS), including Work Programme Providers, utilise the Employability Performance Rating to transparently manage their supply chains, and publish these results regularly on London's Employment and Skills Observatory.
 - London European Programme Management Unit has recommended in their co-financing framework document that all co-financers use the Performance Rating.
 - London Councils are using the Performance Rating with its ESF funded provision.
 - National Offenders Management Unit is using the Performance Rating with its London ESF funded provision.

3 DEFINING HOW THE WORK PACKAGE AIMED TO ACHIEVE ITS OBJECTIVES

Introduction

- 3.1 In this section, we examine *how* the work package intended to achieve its objectives. To do this, we set out the work package’s “logic model” (also known as a “theory of change”). This requires us to understand the sequence of events thought to bring about benefits or change over time. It portrays the chain of reasoning that links investments to results.
- 3.2 Having understood this “theory of change”, evaluation then returns to the theory of change (in the final section) to see if that process is in place, and whether it has been successful in delivering the change hoped for.

The Vision statement

- 3.3 As a whole, the West London Working City Strategy Pathfinder (CSP) has the vision of

“transforming the coherence and effectiveness of the existing structure of services in West London to deliver a sustainable increase in the number of residents in employment and a reduction in child poverty.”
- 3.4 The Employability Performance Rating work package is one of a number of activities designed to deliver that objective. Individual work packages did not have their own vision statements. Instead, each was expected to deliver the above CSP vision.
- 3.5 The vision represents the overarching mission of the work package as a whole. As such, it represents a target itself. Testing the extent to which the Performance Rating system contributed to the delivery of the vision will therefore be the overall finding of the evaluation, and so is discussed in the conclusion.

What was the work package logic model (“theory of change”)?

- 3.6 The logic model describes contextual conditions and rationale, evidence underlying the rationale for the programme, and links inputs to programme activities, outputs, outcomes and longer-term impacts.
- 3.7 We have explained each of these factors in the previous section of this report.

The theory of change was informed by experience in Australia

- 3.8 Since 2001, Australian employment services have been rated by a “star rating” ranging from one to five stars; five stars indicate the highest level of performance. Job Services Australia agencies are rated by the employment ministry every six months (or milestone) on performance, based on placing clients into work and keeping them employed for 13 and/or 26 weeks. The method was initially developed with the assistance of the Universities of Flinders and Adelaide, South Australian Centre for Economic Studies.

- 3.9 Evaluations of the system have indicated that the Star Rating system works. An independent review, completed by Access Economics in February 2002, concluded that the ratings are calculated using a sound approach to performance measurement⁴. An independent review by the Productivity Commission (PC), a statutory authority charged with analyzing and evaluating the delivery of government services. The PC review found that the ‘new framework has many advantages and should be retained.’⁵
- 3.10 However, the OECD and other evaluations have pointed out that the Star Rating system as originally calibrated meant that providers avoided helping those furthest from the labour market.⁶ Changes have resulted, and Star Ratings now calculate actual performance against a comparable average while taking into account local rates of unemployment, job seeker disadvantage and other factors.⁷

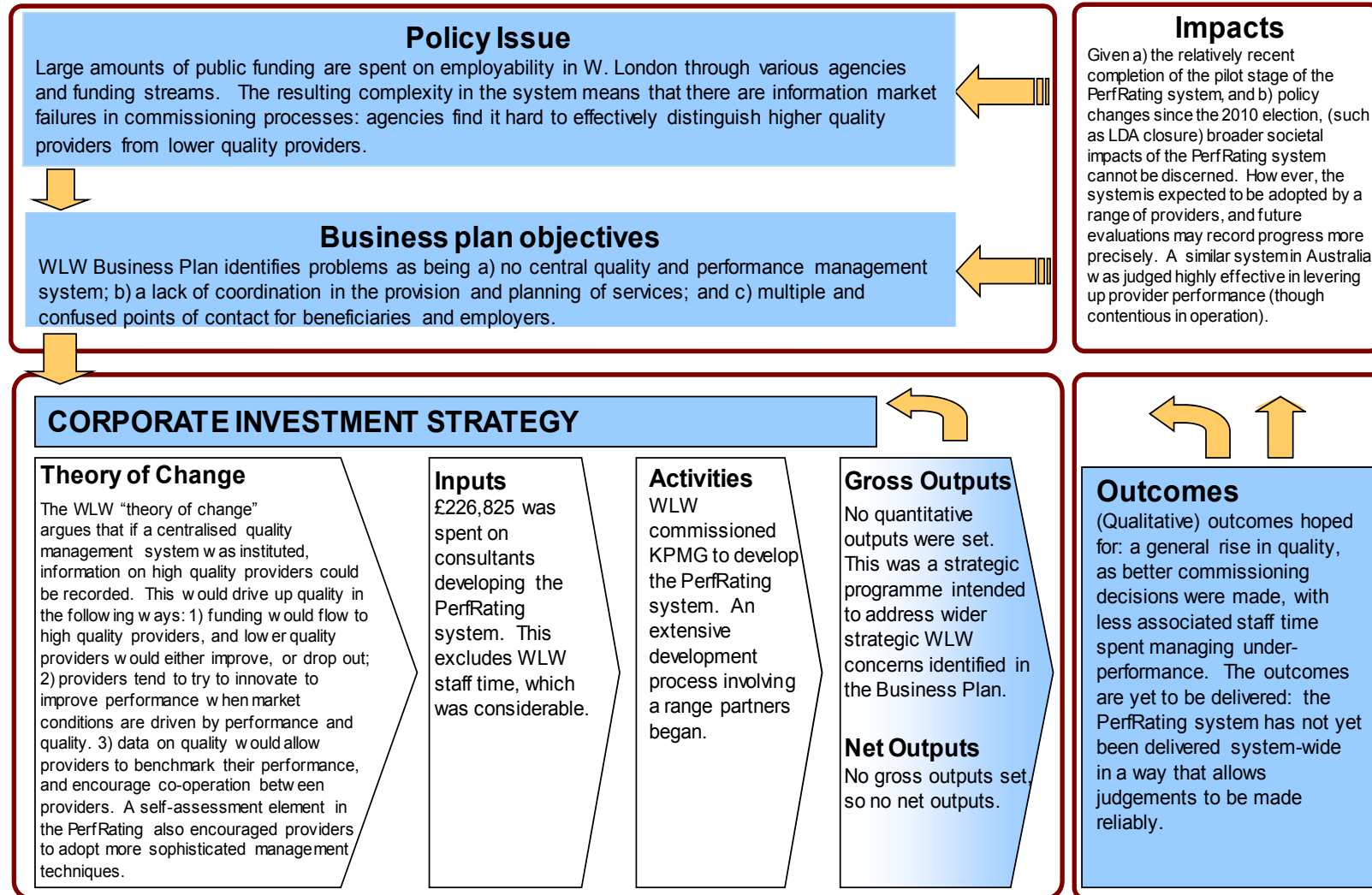
⁴ http://www.workplace.gov.au/NR/rdonlyres/CDB582DF-4937-4B48-881E-6DF647771497/0/Star_Ratings_Fact_Sheet_Feb05.pdf

⁵ Productivity Commission, “Independent Review of the Job Network”, report no 21, pp xxviii, June 2002 quoted Lilley and Hartwich *Paying for Success – How to Make Contracting Out Work in Employment Services*

⁶ See Job Network evaluation Stage three: effectiveness report Evaluation and Programme Performance Branch Labour Market Policy Group Department of Employment and Workplace Relations May 2002; OECD (2005) *Australia – Ageing and Employment Policies*
http://books.google.co.uk/books?id=UKffKRy1WZ0C&pg=PA126&lpg=PA126&dq=australian+star+rating+evaluation+job+network&source=bl&ots=feas-moexk&sig=7v0RXEVYgsz1edtb2noTqbTc0Zo&hl=en&ei=Ax04TZ3rH4y4hAfsqpCsCg&sa=X&oi=book_result&ct=result&resnum=3&ved=0CCgQ6AEwAg#v=onepage&q&f=false

⁷ <http://www.alp.org.au/federal-government/news/address-to-the-national-employment-services-nation/>

Figure 3.1 Work package logic model



4 HAS THE WORK PACKAGE DELIVERED WHAT WAS HOPED?

Introduction

- 4.1 In this section we examine the extent to which this work stream has delivered on the objectives for the work package defined in earlier sections.

Has this work package achieved what it set out to?

- 4.2 Project success is judged in part by progress against its objectives. In short, we ask whether the project did what it was supposed to do.

Reviewing the work package's objectives

- 4.3 It is worth briefly reviewing the project's objectives. In section 3 we explained that the objective of the Performance Rating system was to assist the "marketplace" for providers of employment services function more efficiently, so delivering higher quality services more efficiently. This would be achieved by:
- providing information upon which good quality commissioning decisions could be made, both by public funders, and contractors who were looking to sub-contract to smaller organisations.
 - creating an element of competition around star ratings between providers which would lever up quality through providers' own efforts to improve their ratings;
 - allowing beneficiaries to be directed towards high quality providers by personal advisors; and
 - allowing beneficiaries to select high quality provision from the alternatives on offer.

For full evaluation we need long term, system wide data – but this does not yet exist

- 4.4 Given these objectives, the effectiveness of the Performance Rating system could only be discerned by looking at the performance of those projects using the Performance Rating system across the programme as a whole, over a relatively long period of time. Particular impacts in the case of individual projects might yield evidence, but the real benefits would not become particularly apparent in individual cases: instead, real evidence that the Performance Rating system was working would become apparent across the network of provision as a whole. Ideally, impact would be best judged by using longitudinal data, which has been corrected for the effects of other variables (such as recession). Improvements in quality would be reflected by better achievement of targets, as funding flowed towards higher quality providers, and poor quality providers dropped out of the system.
- 4.5 No such longitudinal data exists for this work package. In many respects, the Performance Rating system has not yet been able to prove its worth in a system-wide way. Although data does exist from the pilot which concluded in July 2009, this is not comparable with the current versions of the Performance Rating system, because the KPIs developed through the pilot period. From April 09- March 10 live versions of the

Performance Rating system were used with the Heathrow programme only and six month “stocktake” data also exists.

Much depends on the future take-up of the Performance Rating system

- 4.6 It is unfortunate that the completion of the live pilot phase of the Performance Rating system has coincided with significant change to the skills and employability landscape. These changes have seen the wind-down and future closure of the LDA, meaning that the Performance Rating system has lost its organisational sponsor, and so its main proponent. Whilst the GLA will take on this sponsorship role, one of the strengths of having WLW leading on the development and delivery of the Performance Rating was that the organisation and personnel had invested in the process of expanding its use. It may be that time and resource limitations at the successor organisation mean that this process might not be run with the same vigour.
- 4.7 The creation of a system of prime contractors means that more responsibility for project delivery will be taken by what are relatively large, well resourced corporates, many of whom have their own management information systems: although the Performance Rating system will be put to use with ESF contracts, it was created too late to be of use during the Primes’ selection of sub-contractors for the Work Programme.

We can use the logic model’s “theory of change” to help evaluate progress

- 4.8 In the absence of long term data on project quality, or certainty on the precise extent to which the project will get taken up, we can turn to the logic model proposed in the previous section. The logic model approach is designed to go some way to solving the problem of demonstrating the linkage between policy actions to outputs on the ground. This process provides some way of addressing a traditional problem in studies of this type – that of attributing causality for positive outcomes to policy interventions. Where outcomes are expected only in the long term (in this case, after the evaluation is complete) they give early indication as to whether predicted changes are happening and therefore whether the intended outcomes are likely to emerge in due course.
- 4.9 In the following section, we therefore examine the evidence regarding the extent to which the Performance Rating system is *likely* to make a real contribution to WLW’s objectives over time. We return to the issue of the logic model in the final section, where we draw together this evidence (and evidence on value for money and Strategic Added Value).

Was there any duplication of provision?

- 4.10 If the Performance Rating system represented duplicated other similar management information packages and processes, its additional benefit would be eroded, so reducing its claim to be a significant part of WLW’s “theory of change”.
- 4.11 Duplication of provision was certainly a risk in this project, and we explored this point carefully.

DWP and the SFA had performance management and tracking systems which fulfilled a similar role to the Performance Rating system

- 4.12 We note that DWP and the Skills Funding Agency (SFA) had performance management systems and associated software packages which fulfilled a similar role to the Performance Rating, or had such packages under development at the same time that Performance Rating was being developed.
- The DWP's system was known as Star Rating, and was based on an Australian system which was considered instrumental in driving up the quality of provision. The British incarnation was developed within DWP, and is a simplified version of the Australian model. Originally, Star Rating was developed for use in Employment Zones. No new software development was undertaken given that the system is based on Excel, although the Star Rating system does contain a number of macros.
 - The SFA's system was known as Framework for Excellence. The Framework for Excellence was the result of a major development programme involving external consultants. The final product was a complex and highly specified system.
- 4.13 The existence of both of these packages was known at the start of the Performance Rating development process. Indeed, the consultants appointed were partly chosen on the basis of their participation in the development of the LSC/SFA Framework for Excellence system. Both WLW and the chosen consultants were also aware of the DWP Star Rating system which was under development at the time of project inception, and its Australian forerunner.
- 4.14 We have heard views consistently expressed that the WLW Performance Rating system was not intended to replace the DWP or SFA systems, or to compete with them. Instead, it intended to fill the "gap in the market" between the DWP Star Rating and SFA Framework for Excellence systems to create a product that could be used by Councils, ESF projects, and sub-regional partnerships. The gap in the market was created firstly, because the DWP system was not designed to work at a sub-contract level, and secondly, because the SFA's "Framework for Excellence" system is said by some of our interviewees to be inflexible in dealing with the needs of smaller scale (often ESF-funded) provision.
- 4.15 Whilst WLW knew that similar systems existed, the question remains whether it was a sensible choice at the time to develop a new system. In an ideal world, of course, funding streams would have similar or even identical reporting requirements, and one performance management system (perhaps with different modules or options) would be rolled out nationally. We therefore investigated the extent to which it would have been sensible to piggy-back on existing DWP or SFA systems.

Evidence suggests that WLW could not sensibly have adopted the Framework for Excellence system in preference to developing Performance Rating

- 4.16 The unanimous response from interviewees experienced in this particular issue has been that the SFA's Framework for Excellence system was highly complex (even over-specified) and very closely tailored to the SFA's organisational concerns. It would therefore have been unsuitable for the types of provision and provider that the WLW/LDA was supervising. We also heard (but have not independently corroborated) that Framework for Excellence

has been the victim of its own complexity, and that getting the data robust enough to be fair has been difficult.

Evidence suggests that WLW could not have straightforwardly adopted the DWP's Star Rating system in preference to developing Performance Rating

- 4.17 Although Star Rating has now been shelved by ministers, who wish to see performance monitored using different methods, the Star Rating still has its advocates within DWP. They point out its merits of relative simplicity, quarterly management capability, and easy data inputting. However, others point out that whilst the DWP system had merits for managing DWP-funded provision (which focused on a narrow range of job outputs) it would have struggled to monitor the breadth of different funding streams were being delivered in West London including provision funded by DWP, ESF, LDA, and SFA. WLW also had a wide range of contract outputs such as level 2 qualifications, voluntary jobs, CV support, and language support. The Star Rating system would not have been able to track these measures and reporting requirements without a major redesign.
- 4.18 It appears, then, that Star Rating could not have been straightforwardly adopted by WLW. However, the question remains whether a modified version of Star Rating could have been used. This might perhaps have seen a new module created, which would have been able to cope with the diversity of provision in West London, but would still have been recognisable as the DWP Star Rating system. We put this point to our interviewees.
- 4.19 Again, the response has been relatively consistent. Interviewees have stated that the Star Rating system was used as the basis for much of the thinking, and DWP confirm that they collaborated in detail with the designers of the Performance Rating system. In that sense, the Performance Rating system in fact *is* a version of the Star Rating system. But we have also heard evidence that the Performance Rating system went considerably further than Star Rating in a number of dimensions. The Performance Rating system included a beneficiary views component; an additional weighting which amplified the points awarded for core contract outputs; and included built-in incentives to over-deliver. This additional functionality was not available from Star Rating.
- 4.20 There are good reasons to think, then, that the Performance Rating system did occupy a “gap in the market” not covered by DWP or SFA systems at the time it was commissioned. The advantages and disadvantages of working with, or modifying, alternative systems were considered. The Performance Rating system does seem to have brought something additional, and was required, given the complexity of the environment that WLW was operating in.
- 4.21 If there is risk of duplication, then, it appears to be that provision is still commissioned through multiple funding streams, each with different management information requirements. But it was not within WLW's gift to impose a single system, and WLW cannot reasonably be criticised for failing to anticipate the change of Government and the shift to the Work Programme.
- 4.22 DWP are not planning to take their internally developed “Star Rating” system further, partly because of shortcomings that ministers perceive related to the system's ranking method, which generates norm-referenced scores against other provision, rather than criterion

referenced scores against fixed criteria. DWP will revert to using plain management data, without further ratings systems or comparisons between providers.

Does the Performance Rating system remain relevant?

Concerns about duplication now also involve how Performance Rating duplicates the internal quality management systems used by the Work Programme Primes

- 4.23 Since the Performance Rating system was commissioned, the landscape has changed considerably. Concerns about duplication now also involve how Performance Rating duplicates the internal quality management systems used by the Work Programme Primes.
- 4.24 Smaller providers hoping to act as sub-contractors to the Primes have found that the Primes either have their own internal performance management systems, or have adopted DWP systems. P3, for example, note that their work sub-contracting to Ingeus is managed using Ingeus' own monthly quality system which works on Red-Amber-Green rankings, whilst their work for A4E's delivery of DWP provision is performance managed using DWP and OFSTED systems, though any future delivery of ESF provision will use Performance Rating. Whilst WLW did not expect the Performance Rating system to replace day-to-day performance management such as the Red-Amber-Green reports, it is the fact that the process of submitting management data looks and feels similar to project providers, and is perceived as creating administrative duplication for providers. P3 seems relatively resigned to this situation, saying that "each funder wants their own version of the information and have their own system for how they want it done". Whilst providers appear to have grown adept at recycling performance data and inputting into different tracking systems, providers note that any simplification would clearly be welcome.

It will be challenging to secure implementation by DWP Prime Contractors without DWP making this a contractual requirement. The DWP have not done this to date

- 4.25 The use of Performance Rating for DWP Provision is dependent on Prime Contractors responding to the requests from London government and the Employment and Skills Board to use the rating, and that they see the business case in implementing it.
- 4.26 We heard from one potential DWP Prime involved with eight community ESF contracts, six of which were co-financed by the DWP, that they intend to use the Performance Rating system for contracts funded by ESF, but they were not planning to take on the Performance Rating system independently for Work Programme funded work, because it effectively duplicated their own systems. "We analyse our own performance internally through their own systems. We are always looking at performance and we do have sophisticated systems."

A potential role for the Performance Rating remains

- 4.27 We have heard further arguments that the Performance Rating system is particularly relevant, for the following reasons.

- The NAO undertook an audit of the Pathways programme which stated that the performance of the supply chain was an area where DWP needed to know more. The Performance Rating can perform this role (although Prime Contractors may have their own systems in place).
- The Merlin standard is DWP's mechanism to ensure that prime contractors treat sub-contractors to an agreed standard. Amongst other things, it requires the use of a code of conduct, and service level agreements. WLW staff argue that the Performance Rating system allows the transparency needed to within large Prime Contractor programmes to see how individual Work Programme contracts are performing. The use of the system would allow sub-contractors to demonstrate that they had been a successful part of a Work Programme work package, even if the overall results of a prime contract had been less than impressive.
- The fact that the Performance Rating system is publically available, whereas Prime Contractors' management information systems will not be, has important ramifications. It means that if prime contractors solely use internal management information systems without making those systems public, the benchmarking function (which saw providers and their supply chains compare their performance) would be lost. This will reduce the opportunities for quality to be driven up through peer effects.
- We have been told that one Prime Contractor has stated that at the moment, Primes are likely to be able to motivate their supply chain through two main mechanisms – that of the “stick” ceasing contracting, and the “carrot” of increasing contracts. The individual thought that there was very little in between these two options, and that the Performance Rating could provide Primes with the necessary leverage because the findings were publicised.

How has the work package expenditure performed against targets?

- 4.28 Expenditure against targets can be used as a proxy measure of activity in some regeneration projects. For example, late starts, commissioning problems and management difficulties can be picked up by measuring spending.
- 4.29 However, with a work package such as this, measures of spending are less useful. As long as consultants are commissioned, they can generally be relied on to invoice on time.
- 4.30 Alterations in spend profile were made during the Performance Rating project, but this was supported by a board-approved change form.

Were changes made from the initial concept? Why?

- 4.31 No fundamental changes were made to the overall concept or objectives of the Performance Rating system through the development process. However, the development process did show that a number of practical issues were dealt with and overcome, as would be expected. We heard evidence that these were dealt with effectively through the development process. For example
- it had been the overall objective of the scheme to get a three-way view of provider quality, with inputs from a) businesses (who had either employed beneficiaries

supplied by contract providers, or who had run work experience programmes) b) DWP advisors/contract managers, and c) beneficiaries themselves. This original objective was modified. It became clear through the development process that there were insufficient numbers of businesses who have an ongoing relationship with providers to provide a good sample size. A small sample size would make data unreliable. Similarly, the rating from Jobcentre Plus advisors was shelved on the advice of DWP. The DWP stated that data was similarly unreliable given sample size. The rating from the beneficiaries was the only measure retained.

- There were also intentions to get a customer satisfaction web tool, but these were not carried forward on the grounds that Development Group members feedback that a paper-based method was easier to incorporate into existing processes.

What could have been improved during the development process?

Interviewees were very impressed by the Performance Rating's development process

- 4.32 The consistency on this point is remarkable. All our interviewees were very impressed by the Performance Rating's development process, and single out WLW staff and KPMG staff for particular praise.
- 4.33 WLW and KPMG staff are praised for running tight meetings with well defined and clearly articulated aims; they involved a range of stakeholders, and listened carefully to their ideas; and explained options and alternatives. The different underlying conceptual approaches to performance rating (such as the relative merits of absolute and comparative performance measures) were discussed thoroughly. The development meetings were seen as "a good use of time"; more than one participant has said that the meetings were enjoyable. Another said that both WLW and KPMG staff were "genuinely interested in what other people knew, keen to learn and to think about other systems".

The precise nature of the relationships between WLW and the LDA was unclear to begin with early in the development process

- 4.34 Funding for projects implementing the Performance Rating system was often from the LDA – so it was essential that the Performance Rating methodology was aligned with the LDA's organisational objectives.
- 4.35 However, early in the project, there were questions about how West London Working's City Strategy Pathfinder status fitted within LDA structures, which created some confusion regarding roles and responsibilities.
- 4.36 Once past the pilot stage, we heard that there was significant buy-in, and LDA contract managers became better involved.

LDA staff reorganisations hampered progress as institutional knowledge was lost

- 4.37 Although unrelated to the work package in question, we understand that there were high levels of staff turnover and some significant staff reorganisations, particularly at the LDA

(although the WLW project lead stayed constant throughout). This required high levels of re-briefing if the Performance Rating was to stay effective.

Investigating whether the development process took too long

- 4.38 If the system had come to fruition sooner – perhaps through an accelerated development process – then some of the benefits of the Performance Rating system may have been more apparent. Critics may argue that piloted and phased delivery is all very well, but timeframes here were so extended that the initial relevance of the project suffered, the context moved, and the work package was left behind.
- 4.39 We tested whether the development process had been too long with relevant interviewees. Interviewees gave a consistent answer. They believed that whilst the development process took a long time, and could possibly have been condensed, this would have needed either a) more intensive input time from project providers and WLW, which many would have found difficult to provide, or b) lower quality consultation with users. In retrospect, it would have been worth accelerating the development process in order to get the Performance Rating system into place. But this could not have been known at the time. At the time, the project was being run collaboratively and very effectively. An accelerated delivery would have damaged the collaborative approach which (as stated earlier) was such an apparently positive aspect of the project.

Were there any unintended impacts or indirect effects resulting from the work package’s activities?

- 4.40 “Action research” is a method of social research in which the research itself is intended to have positive impacts. There is evidence that the process of development had exactly these types of impacts – although in this case the positive impacts have been an inadvertent benefit, rather than an objective in itself.
- 4.41 One view consistently expressed was the very process of developing the Performance Rating system was in itself valuable. The Development Group meetings appear to have been a good way of getting to know other providers in West London, and sharing best practice. We heard evidence that (together with other networking events undertaken by WLW) the Performance Rating development process has proven useful in assisting the larger prime contractors to develop relationships with smaller providers in the area. This was of benefit to both smaller organisations and the larger ones. More importantly, these links would be of significant help to beneficiaries, as they better facilitated the process of passing clients through the system to the most appropriate provision.
- 4.42 One interviewee stated that “the WLW area does seem more joined up than other areas. This has genuine benefits, and helps link everything together, helps [their organisation] as smaller providers are better linked to bigger ones. The Work Programme commissioning process might be better [in the WLW area] than elsewhere due to these relationships being built up.”
- 4.43 We did not hear evidence of any unintended negative consequences or effects of the Performance Rating work package.

What has worked well and not so well?

Worked well: the development process

- 4.44 As discussed earlier in this section, WLW's development of the Performance Rating system seems to have been an excellent example of collaborative development. As one participant to the project put it, those managing the project (from both the consultant and WLW side) were "willing to listen, willing to talk and willing to discuss. There is nothing difficult about that but it is amazing how often we manage not to do it".

Worked well: driving up quality

- 4.45 We heard strong support for the system, and a broad belief that it would contribute to quality outcomes if used more widely in future.
- 4.46 One interviewee stated that "The employment performance rating I think is an incredibly good piece of work, which really should be taken on more widely." The same interviewee stated that the some kind of measurement system was "vital", and that this tool gives an indication on who is providing effective services whereas it might be "guesswork" in other sub-regions.
- 4.47 The more ambitious providers were keen to use the system. As they saw it, it would be useful to demonstrate quality to funders, and to incentivise staff. One provider interviewee thought that from his perspective "you need to know how you are doing compared to others. We can use this sort of information to motivate staff and understand our own provision...I am a believer in healthy competition". One interviewee expressed the hope that the more competent providers would be able to use the Performance Rating system as a way of improving their own performance, and display that quality to funders, and so win a bigger market share.
- 4.48 Separately, WLW has also received email enquiries from providers who wish to know whether they might use the Performance Rating system to generate their own ranking – which is presumably intended to be of use in attracting funding from (say) ESF or interest from Prime Contractors.
- 4.49 Although there was broad support for the system, one interviewee dissented. One interviewee from a Local Authority thought that the performance rating system was overly bureaucratic and not particularly helpful.

Worked well: managing the supply chain

- 4.50 We heard evidence that many delivery partners in the third sector are finding it hard to manage their sub-contractors. We heard evidence that the Performance Rating system
- Helps lead partners see how their sub-contractors are doing. For example, with the ESF Youth projects, all the lead partners have sub-contractors. With Performance Rating in place, the LDA gets a view of the quality of the supply chain; also lead partners can use it to lever up the project quality of their contractors.
 - Lets sub-contractors see that parts of the contract are not on their own especially valuable (some targets are not "prime outputs") so that they focus on the right, most crucial elements of the targets.

Worked less well: the Self Assessment element is valuable, but providers need more support in using it well

- 4.51 One interviewee felt that a significant benefit of the Performance Rating system was the way in which it allowed the smaller providers to undertake a self assessment process. Self assessment systems are used by the larger providers to identify areas of improvement, and this was seen as a way of levelling the playing field between smaller and larger providers by introducing higher quality management techniques to smaller, often community-based organisations. One interviewee stated that “Getting [projects] to think about their practice and performance was a very positive thing.”
- 4.52 The self assessment element of the Performance Rating system accounts for 9% of the total marks awarded, and so counts towards the number of stars that a contract holder is awarded. Clearly, there is something of a moral hazard here, given that the contract holder has the means of boosting their own score.
- 4.53 This point did not escape work programme bidders. The decision to retain a self-assessment element was made because it was felt that a) it would encourage a reflective, self-critical approach to delivery on the part of providers, and b) that without the self assessment process being included as part of the assessment, it would never get done. Detailed and thoughtful templates were provided to assist the self assessment process.
- 4.54 Two main problems have been experienced with the self-assessment section. Firstly, we understand that the self-assessment process has been very uneven: some providers have done a good quality, careful piece of work, whilst others have made little or no effort.
- 4.55 Secondly, the point about moral hazard does seem to be borne out in practice. We understand that in many cases, draft self-assessment ratings seemed inversely related to overall performance - suggesting that projects which have been weak deliverers are tempted to “claw back” marks by awarding themselves high self-assessment marks, or at the very least a lack of awareness about how other providers were performing.
- 4.56 We heard evidence that self assessment could be a very powerful component in helping projects improve the quality of provision. We understand the value of self-assessment, but think that the precise way that this element works might be reconsidered. There does appear to be something wrong with having projects able to influence their own rating. An honest, frank self assessment is important. The chances of this happening are reduced if providers believe that their scores are connected to this self-assessment.
- 4.57 These were difficult choices that were not lightly made, and we respect the conclusions that the Development Group and Project Board came to. Even so, we suggest that WLW (or other future users of the Performance Rating system) might consider disconnecting the self-assessment from the overall generation of rankings, but making a high quality self-assessment a mandatory part of contracts. The quality of self-assessment might also benefit from the delivery of a small number of day training sessions, so that providers get the best from the work that has gone into the self-assessment template.
- 4.58 We note that providers funded through the Work Programme will now be required to undergo self assessment processes, and will be subject to light-touch inspection from the DWP’s Provider Assurance Team. A single self assessment tool will be created for DWP

provision.⁸ However, non-Work Programme provision (such as ESF) does not have mandatory self-inspection (nor external inspection), and so the maintenance of the self-inspection element through the Performance Rating system remains valuable.

Worked less well: the system does not operate well on a quarterly or six-month timeframe

- 4.59 Ideally, a Performance Management system would allow ongoing monitoring of subcontractor performance to take place. This would allow contract managers to identify when a sub-contractor’s project was going off course, in order that early corrective action could be taken.
- 4.60 However, the system has been designed as a system which operates on a year-by-year basis – in order that this matches the typical contract period.
- 4.61 The WLW Associate Director has successfully created a workaround to enable like for like publication of a six month stocktake of the Performance Rating, but reports that the system does not readily allow this to take place.

How (if at all) does the work package link to other WLW activities and objectives?

- 4.62 The WLW Business Plan identified this theme of improved information flow as being one of its main objectives. In this sense, the Performance Rating work package is one of WLW “core packages”. WLW’s objective was to provide an integrated package that would knit together and streamline employment and skills provision in West London. There were a number of elements in this programme.
 - The Investment and Target Group was intended to encourage joint working between the organisations represented leading to aligned or joint planning, commissioning and investment in services targeted to reduce worklessness in West London. The desire was to improve coordination, so that services fill gaps and duplication is reduced. The group also aims to disseminate management information and work from a common information set.
 - Building Employability Pathways is designed to support West London residents move from learning into sustainable employment. Common forms of activity include Colleges developing stronger links with Jobcentre Plus and other relevant job brokerage organisations so that their learners have great access to vacancies; building links with other employment and skills provision, and local employers; and introducing employability skills into the college’s practices.
 - The Provider Confederation sought to build a network of providers of employment services. The Confederation’s objectives are to reduce duplication by ensuring that providers are better aware of nearby provision; build providers' capacity by using events

⁸ ‘Ofsted inspection to end for DWP Provision’ Friday, August 13th, 2010 <http://www.carleyconsult.co.uk/news/?p=445>

and e-bulletins to spread best practice; and improve personalisation of services by encouraging providers to refer residents onto services which best meets their needs.

- 4.63 There is good evidence to indicate that the Performance Rating was a well integrated part of this overall effort, and was one which would, on implementation, have added value and synergy across these processes.
- Information from the Performance Rating system would have bolstered the Directory. It was always an objective to publically publish the star ratings generated by the Performance Rating system, and one of the ways that this would have taken place was through the Directory. Prime contractors and sub-contractors would have had better information about the performance of provision delivered by different organisations, and would have been able to use this information to come to better informed partnering arrangements and sub-contracting agreements.
 - Information from the Performance Rating system would have bolstered the Investment and Target group, allowing the members of this group to make better informed investment and targeting decisions.
 - Information from the Performance Rating system would have bolstered the Confederation – how providers can manage their supply chains and chose people to work with.
- 4.64 Information from the Performance Rating system would have benefited beneficiaries directly. The star ratings would have been available to personal advisors, who would have been able to use the ratings to point clients towards higher quality provision. Over time, the flow of clients to higher quality organisations would have generated internal pressures that would have forced lower quality providers to either drive up quality or to exit the market.

5 EVALUATING THE STRATEGIC ADDED VALUE ACHIEVED

Introduction

- 5.1 In this section, we examine the extent to which Strategic Added Value (SAV) has been delivered by the work package.
- 5.2 Originally, the idea was created in order to better evaluate the contribution made by Regional Development Agencies, but it is the case that the concept applies equally well to the City Strategy Pathfinders.

What is Strategic Added Value (SAV)?

- 5.3 Strategic Added Value (SAV) has been defined as the “non work package activities that do not produce traditional outputs” such as the ‘catalytic’ and ‘influencing’ role.
- 5.4 The idea of Strategic Added Value is to ensure that there is a way of capturing the benefits of strategic programmes such as City Strategy Pathfinders that go beyond the immediately obvious outputs and outcomes. This recognises that the contribution of City Strategy Pathfinders can come through their work package and programme spend but also through their influence on partners and stakeholders’ behaviour and performance.

How we evaluated SAV

- 5.5 We set up a SAV evaluation framework. Table 7.1 shows the five different categories of SAV and shows the types of evidence that we looked for in order to investigate whether WLW generating this form of added value in this work package. The table also shows the average Likert scale responses across the eight interviewees we talked to in depth on the Performance Rating system (we talked to eight other stakeholders about the Performance Rating system, but time pressures meant that these Likert Scales were only explored with the eight in-depth interviewees).

Findings on SAV

- 5.6 Interviewees were generally positive about the way that WLW’s Performance Rating system development phase had added strategic value. The area where interviewees felt that the work package had most added value was in the area of strategic leadership and catalyst (“Improving the capacity of partners and stakeholders to realise the potential for improved performance; leading partners towards those objectives”); the area it was lowest was in the area of leverage (“levering funding and other resources from partners and stakeholders in support of shared objectives”).
- 5.7 Our questions concentrated on the SAV added so far – and so interviewees’ answers referred to the development phase of the Performance Rating system. As the Performance Rating system is rolled out and used in practice, further Strategic Added Value benefits may be experienced.

Likert scales and analysis can be found in the table below.

Table 5.1 Strategic Added Value Likert statements

SAV category	Understanding required	With this work package's development phase, WLW has successfully...	Likert score by stakeholder (Strongly disagree (5)/ disagree(4)/ neither (3)/ agree (2)/ strongly agree (1))	Analysis and further detail
Strategic leadership & catalyst	Improving the capacity of partners and stakeholders to realise the potential for improved performance; leading partners towards those objectives	...articulated the problems that project clients were facing in commissioning and monitoring successful projects	1.4	<p>This was the strongest agreement of the different SAV factors. Stakeholders felt that, with the Performance Rating project, WLW had evidenced an understanding (and had actively sought to understand) commissioning and performance management issues.</p> <p>One stakeholder interviewee stated that "it was clear that there was a raft of public organisations getting money without any quality monitoring. This was trying to plug that gap. It is a tool that can commission new interventions and benchmark themselves against one another."</p> <p>One Prime saw this as one of the main benefits of their work with the Performance Rating development group - they were able to communicate to smaller organisations how they contracted, and better understand the challenges of working with the third sector.</p>
		...catalysed cross-organisation action to solve the problems experienced	1.7	<p>One interviewee stated that a wide range of "partners and providers came on board. It was all very well managed" whilst another noted the "inclusivity" of arrangements that encouraged cross-organisational action.</p> <p>There were doubts about the involvement of the LSC/ SFA, however, but this was not seen as a fault of the management of the Performance Rating development process.</p>
Strategic influence	Generating cross-regional partnerships of mutual benefit to the growth prospects of each participating region; Achieving alignment and inter-locking of the priorities and investment plans of partners with the RES and with other partners;	... encouraged partners to commit to shared strategic objectives	1.8	<p>In line with comments above, interviewees noted that a wide range of partners and providers came on board during the development process.</p> <p>However, one interviewee did note that this work did not involve partners committing to alter their strategic objectives - it was simply a case of encouraging joint action from partners with objectives that were already strategically aligned.</p>
		...encouraged partners to work in a compatible way, and to share information on project quality	1.9	<p>One interviewee stated that the development process encouraged partners to work in a compatible way, and that the Performance Rating system itself would have encouraged more cross partner working</p>
Leverage	Levered funding and other resources from partners and stakeholders in support of shared objectives	.. levered funding and other resources from partners and stakeholders in support of WLW objectives	2.5	<p>This was an area where interviewees felt that WLW had been least successful, although some noted that considerable time had been levered in from a wide range of partners attending the development board.</p>

SAV category	Understanding required	With this work package's development phase, WLW has successfully...	Likert score by stakeholder (Strongly disagree (5)/ disagree(4)/ neither (3)/ agree (2)/ strongly agree (1))	Analysis and further detail
Synergy	Reduced duplication of service provision and driven up provision quality	...used expertise to improve information exchange and knowledge transfer and coordination between partners	1.5	Interviewees thought that this was the second strongest area for WLW. One interviewee felt that these synergy benefits "will increase as it runs more widely with London Councils and NOMS. The system has already helped procurement process on ESF Youth - easy to find people to talk to, and we know more about people in the LSC...[these benefits] will only increase as it gets used".
Engagement	Scaling up of work packages and programmes to beneficial levels that achieve scale economies and provide for critical mass in securing benefits;	...engaged partners scaled up of work packages and programmes to achieve scale economies and provide for critical mass in securing benefits	2.3	One interviewee stated that the Performance Rating system had potential to create scale economies in the commissioning process, "especially if used through the new Work Programme. But hasn't happened yet". One interviewee stated that he was not working together more as a result - "pooling is happening, but not as a result of the Performance Rating system."

Source: SAV categories Quoted p20 DTI document *England's Regional Development Agencies: Developing A Methodology And Evaluation Framework*; Adapted from the Guidance Note on SAV prepared by SEEDA for the RDA's Performance Management Group and from the ECOTEC Scoping Study for *emda* (2004)

6 EVALUATING VALUE FOR MONEY

Introduction

- 6.1 In this section, we evaluate the extent to which the work package represented value for money.
- 6.2 No formal evaluation of value for money will be possible, as there is a lack of quantitative output data to compare against cost and spending information. However, we can evaluate value for money more qualitatively.

Did the work package represent value for money?

- 6.3 As we showed in section 3, £226,825 including VAT was spent on this project. This money was paid to the contractor, KPMG. Considerable amounts of WLW staff time were also spent on this project. This staff time is not included in the sum quoted above.

Value for money will depend on how the system is used in future

- 6.4 It is too soon to say whether this project represents value for money. This is because the project has, at the time of writing, not yet been fully rolled out across a range of employment and skills provision. As a live (pilot) system, it has been used in the Gateway Heathrow 2012 LDA ESF project, with twelve LDA prime contracts (and an additional 32 subcontractors, and with London Council's ESF programme. Whilst there is evidence that further roll-out will take place (as we have detailed earlier, there is considerable buy-in to the concept of the Performance Rating system), any judgement on the value of the project depends entirely on whether the Performance Rating system is used widely. In short
- If the Performance Rating system is shelved (whether due to organisational change, a shift in political priorities, or for technical reasons) then clearly the £227k spent on the project will prove to be very poor value for money.
 - If the Performance Rating system is widely adopted, and proves valuable in levering up the quality of provision and the quality of management decisions, then the scheme could prove to be good value. As one LDA interviewee stated, "the problem is that we spend a huge amount of time managing underperformance...if the system can fix that problem, [the Performance Rating system] will be worth its weight in gold." WLW points out that the annual spend on employability in West London is in the region of £60m pa, and London-wide £200m pa is spent. If the Performance Rating system was to yield even small efficiencies and synergies in this provision, then the level of investment on the project would be worthwhile.
- 6.5 As we have shown in paragraph 2.23, the Performance Rating system has already been implemented London-wide (beyond the immediate borders of the WLW area), through the use of the rating by twelve of the LDA's major learning and skills projects. Its use in London has been further bolstered through the endorsement of the London European Programme Management Unit (EPMU), which will require use of the Performance Rating system in all relevant ESF projects in London. The London Skills and Employment Board is encouraging all of London's employment and skills services to use the Performance

Rating system in tracking their performance and that of their supply chain providers. London Councils and National Offenders Management Services are just starting to use the system from October 10. All LDA projects in 2011/12 are also committed to use the system, as presumably would projects run by any body running successor programmes to the LDA such as the GLA.

The contract to develop the system was let through open contract

- 6.6 There is then the question of whether the project's outputs could have been achieved for less money. KPMG were chosen following an open tendering process. WLW staff noted that KPMG had comparatively high day rates, but that KPMG had added considerable value, and had worked in an obliging and collaborative way.

7 SUMMARISING THE LESSONS LEARNT AND DRAWING CONCLUSIONS

Introduction

7.1 In this section, we draw some overall conclusions from our evaluation.

Summary of findings

- 7.2 **The logic model’s “theory of change” remains effective as a piece of logic. We see the strategic rationale for the development of the Performance Rating system perfectly:** that, in a very crowded “marketplace” of providers of employability services in West London, there needed to be a way of funding the best providers and avoiding funding the worst, and allowing providers to benchmark quality. Without the Performance Rating system, there were simply too many ways that poor quality providers could hide poor delivery, even if existing management information systems were working properly. Experience in Australia suggests that a system like the Performance Rating system can make an important contribution to driving up quality, if the incentive structure created through the star rating system is well aligned to the desired outcomes.
- 7.3 As such, **the Performance Rating system was very sensible response to a classic “information” market failure.** It was and is an excellent example of how proper knowledge management techniques could potentially improve public sector procurement, and so improve the quality of service provided to low skilled and often deprived individuals. As well as providing better commissioning information, the system also sought to generate a number of subsidiary benefits, such as introducing improved management techniques within providers organisations (in particular, via the thoughtful self-assessment template provided as part of the work package).
- 7.4 **As well as being strategically sensible, the system was well conceived and well delivered in practice.** Performance Rating system did not, at the time of inception, duplicate existing provision, and instead seems to have successfully occupied a “gap in the market” between the DWP’s (now shelved) Star Rating system and the SFA’s apparently highly complex Framework for Excellence system. By common consent, the Performance Rating system was developed in textbook fashion, with genuine collaboration between the consultants developing the system, the WLW client group, and the development group made up of users and stakeholders. Our interviewees suggested that the Performance Rating system generated considerable Strategic Added Value for WLW during the development.
- 7.5 **However, the way that the logic model’s “theory of change” plays out in the real world has yet to be seen. Real evidence that the Performance Rating system was working would need to be seen across the network of provision as a whole.** Improvements in quality would be reflected by better achievement of targets, as funding flowed towards higher quality providers, and poor quality providers dropped out of the system. However, **longitudinal data is not available for this work package.** This is because it is too soon for this to be available.

- 7.6 **There are questions about whether the Performance Rating system will be given the chance to prove itself in practice. The problem is that circumstances have changed.** Whilst the logic underlying the Performance Rating system remains entirely sensible, the medium and long-term future of the Performance Rating system is now questionable. This is due to events that could not have been reasonably foreseen by WLW staff or the WLW Board. As one interviewee from a provider organisation said, the Performance Rating system “has been a victim of circumstance and timing.” This is for the following reasons.
- **The Performance Rating system is in the process of losing its sponsoring organisation** (the LDA). Whilst the GLA will sponsor the Performance Rating system in the absence of the LDA, the GLA has not made the same organisational investment in the system. The risk must be that without a very active sponsoring organisation, the system may fall into disuse over the medium term.
 - **Most significantly, the Work Programme represents a very significant change in the marketplace for employability services.** In effect, the onus of programme management is now being placed on the Work Programme’s Prime Contractors rather than public sector funders. In spite of the Mayor’s political influence, it will be very challenging to get Prime Contractors to adopt the system without DWP contractual requirement.
- 7.7 **It is not possible to make quantitative calculations of value for money of the Performance Rating system.** One reason is that, as just explained, the work package has not been run across a wide spread of provision. The second reason is that there are no specific output targets to measure spend against (this project was about a strategic intervention to deliver improved quality). **We therefore looked at value for money more qualitatively.** £227k was spent on the project, which was subject to open tender. If the Performance Rating system is shelved, then clearly this will prove to be very poor value for money. If the Performance Rating system is widely adopted, and proves valuable in levering up the quality of provision and the quality of management decisions, then the scheme could prove to be very good value. As one LDA interviewee stated, “the problem is that we spend a huge amount of time managing underperformance...if the system can fix that problem, [the Performance Rating system] will be worth its weight in gold.”
- 7.8 These changes which have affected the marketplace in employability have their own rationale, advantages and disadvantages, which need not detain us here. **But as far as the narrow considerations of the Performance Rating system itself are concerned, these changes are a great pity. In many respects the Performance Rating system has not yet been able to prove its worth in a system-wide way – although, having interviewed stakeholders and tested the sense of the logic model we are persuaded that its success would be likely, were it to be given the chance.**
- 7.9 Equally, though, **it may be that the shift to the Work Programme means that Prime Contractors’ own management information systems could ultimately perform a very similar – although not identical - function to that envisaged for the Performance Rating system.** This would suggest that the performance management ideas which underpinned the Performance Rating system would survive, but in different form. Although the Performance Rating system cannot claim to have inspired Prime Contractors’ internal

management tools in any way, provision to beneficiaries would be improved. This, after all, is the central consideration. However, the lack of transparency around Prime Contractors' data could mean that the opportunity offered by the Performance Rating system - of providers using the system to benchmark provision between themselves – would be lost.

- 7.10 **The positive reaction of our interviewees suggests that future deliverers of employment services, both inside and outside the Work Programme, should give serious consideration to using the Performance Rating system.** The results of the development period have been a system which has widespread support, and whilst some knotty (and really quite insoluble) philosophical issues remain (about, say, the merits of relative and absolute measurements of success), the system could still make a genuine contribution to delivering quality employment outcomes. The Performance Rating system can provide valuable public information on the performance of supply chains; help to deliver Merlin standards; and provide Primes with the helpful leverage with sub-contractors because the findings were publicised.
- 7.11 **The Performance Rating system would be particularly appropriate to use for ESF funded provision.** Whilst there will be use of the Performance Rating system in the very substantial ESF provision in London (and other applications), it is also of relevance in other European countries, where different national systems and (in some places) relatively higher levels of ESF provision might allow extended use of the system.
- 7.12 **Adopting the system could be relatively straightforward.** The LDA Employment and Skills Directorate would be willing to share a non-protected version of the performance rating calculator excel spreadsheet with public sector providers. The London Development Agency, as accountable body for West London Working, holds the Intellectual Property Rights for the Performance Rating system, but would encourage other public sector providers to adopt it.

Lessons that inform future approaches

- 7.13 If funders were to adopt the system, it would be helpful to make some remarks on what they would need to consider. We address these points below.

There is a debate about whether relative or absolute measurements of provider success are superior

- 7.14 Future users of the Performance Rating system need to be aware of the debate between the merits of relative or absolute measurements of provider success. To be clear:
- relative measures of provider success compared one provider's activity against other providers. This approach was taken by the DWP's Star Rating system. The advantages of such a method are that it identifies the best performing providers in a group, and does place undue emphasis on the realism of the contract targets (in an environment where accurate targets are often very hard to construct). The disadvantages are (as ministers believe) this method means that this does not identify instances when all provision is of a generally poor quality.
 - Absolute measures of success compare a provider's success against the contract targets. This is the method adopted by the Performance Rating system. The

advantages of such a method are that, firstly, this identifies the projects delivering the greatest benefits against the strategic objectives set in the contract, and also incentivises sensible contract setting. The disadvantages are that projects' performance can be dramatically improved if projects are able to renegotiate their contract targets downwards, as has happened in some instances; and that if a contract is insufficiently challenging, there is no incentive to improve.

- 7.15 After detailed discussion of the pros and cons of each method, the development group decided to go with an “absolute against contract” method. It transpires that this is the method seen as superior by ministers. But future adopters of the system need to be aware of the debates around the relative merits of each method, and the implications for contract setting.

Some questions about the precise workings of the system remain

- 7.16 Future users of the Performance Rating system need to be aware of outstanding questions about the precise workings of the system. These are as follows.
- There is ongoing discussion between WLW and providers about whether the Performance Rating system should use an annual rating (where star ratings would run for the length of each contract period, and be reset with agreed profiles at the beginning of each new financial year) or a lifetime rating (which would be cumulative over each year of delivery). There are reasonable arguments to be presented on both ways of running the system. Providers in favour of lifetime ratings suggest that this method would provide a better longer-term perspective on overall performance; WLW staff, who favour the existing annual system, think that this creates risk that ratings do not reflect where project providers have agreed annual contract increases or decreases due to external factors outside of their control.
 - There appears to be a need to develop a better calculator which links the Performance Rating system to the performance returns made by providers. This would reduce the need for any data re-keying. We understand that London Councils have successfully merged these data input sheets into one, so this appears to have been largely overcome.
 - There is a legitimate complaint made by providers that their ratings do not take account of late issue of grant funding agreements. It is fair to say that late starts are endemic across nearly all regeneration schemes we have evaluated around the country, due to delays in project procurement and approval. However, contract outputs are usually not amended to reflect that late start, and the Performance Rating system penalises failure to comply with target outputs. Providers' complaints are legitimate here; we hope that the sense of urgency that the Performance Rating system generates within provider groups can be transmitted back to contract funders by providers themselves. If outputs were defrayed pro-rata with the contract time lost due to funding delays, it might focus funders' efforts to get Grant Agreements out on time.
 - Learner views form part of the assessment. There are problems with placing too much reliance on learner views (learners may have little English; forms can be a poor way of capturing information; learners often have no benchmark upon which to make sensible judgements of quality; and learners can be influenced to fill in forms in a particular way).

9% of marks in the Performance Rating system were allocated to learner views. Good quality external assessment is to be preferred, but it seems to us that the level of emphasis here was broadly right.

The running and management of the system will need co-ordination

- 7.17 If the Performance Rating system was adopted elsewhere, it would need a co-ordinator. That person’s role would be to provide a point of contact for training and queries on using the system; pulling together the performance rating scores; making sure that all information is received every April; ensuring that the self assessment process was being done properly; and spot checking data. If Work Programme Prime Contractors were to use the system, they would need to run it for their sub-consultants.

Future roll out would benefit from targeted publicity

- 7.18 One interviewee suggested that if the Performance Rating system was going to be rolled out further, further publicity about how the system could be used would be helpful. Specifically, it would be important to
- make sure Next Step and Matrix IAG providers are able to use the ratings provided by the system to advise clients; and
 - speak to the Prime Contractors. Primes would be able to potentially use the future ratings generated by small providers through involvement in (for example) ESF provision to better manage their supply chains.

Self assessment work could be extended

- 7.19 The self assessment templates and thinking could be extended. We make this suggestion because we heard evidence that there is a real gap in quality management for some projects, particularly those community-level (often ESF) projects funded in areas not subject to OFSTED or regular inspection.
- 7.20 The Performance Rating system could be analysed to flag up areas of particular concern: one interviewee felt that it would be instructive to identify those projects with poor “real” contract performance and a high, very positive self assessment. As one interviewee said, “if projects are not brave enough to be honest with themselves, there is no point in self assessment”. We heard that this process would tend to identify those projects with management which may be ignoring problems, and those projects which were not able to use the opportunity that self assessment offered for improvement. Then steps can be taken – perhaps pairing good organisations with weak organisations, or offering additional management training where required.

Self assessment ratings could be treated differently by the system

- 7.21 Due to the potential moral hazard of allowing a self-assessment to affect a project’s overall rating, we suggest that WLW (or other future users of the Performance Rating system) might consider disconnecting the self-assessment from the overall generation of rankings. A high quality self-assessment would remain a mandatory part of contracts to ensure it was carried out.

Future funders may wish to be more aggressive in their use of the Performance Rating system – but this would need careful consideration

- 7.22 The Australian model (which was the forerunner of the DWP Star Rating, and therefore of the Performance Rating system) required that a minimum star rating be reached if a projects to access public funding. This significantly raised the stakes for projects, and compelled them to reach a certain number of “stars” if they were to survive; many did not. Whilst a blunt tool, we understand (though have not separately researched) that this approach did seem to raise the quality of provision, although it did tend to reward the “cherry picking” of easier-to-help candidates. Future providers may wish to investigate this more aggressive approach to integrating performance rating with the commissioning process: one interviewee noted that this approach might create both “big problems, and big successes”.