Independent Review of Framework for Excellence

LSC National Office

Final Report

Date: November 2008
CONTENTS

CONTENTS................................................................................................................................. I
GLOSSARY ................................................................................................................................... III
EXECUTIVE SUMMARY .............................................................................................................. 1
1 INTRODUCTION AND APPROACH ....................................................................................... 3
   1.1 Introduction and background.......................................................................................... 3
   1.2 Study approach ............................................................................................................ 3
2 OVERVIEW OF FFE .................................................................................................................. 6
   2.1 Background .................................................................................................................. 6
   2.2 Current coverage – user groups and their requirements ............................................ 7
   2.3 Overview of FFE – comparative review ...................................................................... 7
   2.4 Overview of FFE – stakeholder perspectives ............................................................... 11
   2.5 Issues arising from stakeholder consultation ............................................................ 11
   2.6 Section summary ........................................................................................................ 16
3 CROSS CUTTING ISSUES ........................................................................................................ 18
   3.1 Introduction ................................................................................................................ 18
   3.2 Contextualisation ....................................................................................................... 18
   3.3 Use of surveys ............................................................................................................ 19
   3.4 Composite measures ................................................................................................. 22
   3.5 Combination rules ....................................................................................................... 23
   3.6 De minimis criteria and exemptions .......................................................................... 24
   3.7 Balance of weighting ................................................................................................. 25
   3.8 Summary .................................................................................................................... 26
4 RESPONSIVENESS DIMENSION ............................................................................................. 29
   4.1 Key Performance Area: Responsiveness to Learners .................................................. 29
   4.2 Key Performance Area: Responsiveness to Employers ............................................. 31
5 EFFECTIVENESS DIMENSION ............................................................................................. 34
   5.1 Key Performance Area: Quality of Outcomes ............................................................ 34
   5.2 Key Performance Area: Quality of Provision ............................................................. 36
6 FINANCIAL DIMENSION ......................................................................................................... 38
   6.1 Key Performance Area: Financial Health ................................................................. 38
   6.2 Key Performance Area: Financial Control ................................................................. 38
   6.3 Key Performance Area: Use of Resources ................................................................. 38
   6.4 General issues concerning the financial dimension .................................................. 39
7 CONCLUSIONS AND RECOMMENDATIONS ...................................................................... 40
   7.1 Overview .................................................................................................................... 40
   7.2 Issues to be considered and/or addressed ................................................................. 41
7.3 The overall 'health' of FiE ................................................................. 46
ANNEX A – KEY RESEARCH QUESTIONS ................................................. 47
ANNEX B – PRINCIPLES OF DESIGN, DEVELOPMENT AND IMPLEMENTATION ........... 48
GLOSSARY

ALP – Association of Learning Providers
AoC – Association of Colleges
CVA – Contextual Value Added
DCLG – Department of Communities and Local Government
DCMS – Department for Culture, Media and Sport
DIUS – Department for Innovation, Universities and Skills
FIE – Framework for Excellence
HEFCE – Higher Education Funding Council for England
IAG – Information, Advice and Guidance
KPA – Key Performance Area
LSC – Learning and Skills Council
NAO – National Audit Office
Ofsted – Office for Standards in Education
OPR – Overall Performance Rating
PCDL – Personal and Community Development learning
PI – Performance Indicators
PCDL - Personal and Community Development learning
QAA – Quality Assurance Agency
QIA – Quality Improvement Agency now the Learning and Skills Improvement Service (LSIS)
RSL – Registered Social Landlords
TQS – Training Quality Standard
EXECUTIVE SUMMARY

1. In 2008, the Framework for Excellence (FfE) Policy Committee (comprising LSC, DIUS, Ofsted and LSIS) agreed that an Independent Review of FfE should be conducted as a ‘healthcheck’ of the FfE arrangements. GHK was commissioned by the LSC National Office to conduct the Review.

2. The Review was separate from, though linked to, other FfE-wide assessments and analyses, including: the long-term evaluation of FfE (being conducted by York Consulting); the contextualisation assessment (recently conducted by RCU); and the assessment of users’ requirements for FfE published information (also being undertaken by GHK).

3. The Review was informed initially by a series of key research questions, which can be found in Annex A of the report. As the Review progressed, the questions became condensed into three main areas for consideration: Is FfE measuring the right areas? Is FfE measuring these areas in the right way? Are the data and results being interpreted in the right way? (See Section 1.1)

4. A multi-method approach was taken in order to achieve the main aims of the study: stakeholder consultation; a policy and document review; a review of evidence on learner and employer choice; a comparative review of performance management/measurement frameworks, and; a quantitative review of FfE. (Section 1.2)

5. The report is structured as follows:

   - **Section 2** provides an overview of the FfE, detailing user groups and their requirements, and goes on to highlight some of the issues raised during the comparative review, and stakeholder consultation.
   - **Section 3** introduces and reviews some of the cross-cutting themes which relate to FfE and performance frameworks more generally.
   - **Section 4** provides a detailed review of the key elements comprising the Responsiveness dimension of the framework.
   - **Section 5** presents a review of the Effectiveness dimension and the composite elements.
   - **Section 6** details and reviews the core components of the Financial dimension.
   - **Section 7** presents a summary and sets out key conclusions and recommendations.
   - **Annex A** presents the key questions underpinning the Independent Review.
   - **Annex B** presents the underpinning FfE principles of design, development and implementation.

6. The authors conclude that the FfE has been successful in meeting its stated aims and objectives and addressing the three main areas for consideration in this Review. The FfE is:
• measuring the right areas in order to present an accurate picture of provider quality and performance;

• measuring things in the right way: our Comparative Review found that FfE compares favourably in relation to other existing systems;

• generally being interpreted in the most appropriate way.

7. The evidence suggests that the level of consultation, research, discussion and debate, as well as technical knowledge employed, has resulted in a Framework which offers a relatively robust picture of provider quality and performance (Sections 7.1 and 7.3). The study identified a set of issues of which the LSC need to be aware, and some which also necessitate action to resolve them (see Section 7.2).

8. The authors recommend that:

1. The LSC and its partners should continue to monitor the use of surveys to generate scores, where they are based on incomplete samples because they present an inherent issue of sampling error and, correspondingly, of reliability and robustness.

2. FfE should provide guidance in the interpretation of OPRs where providers are exempted from an entire KPA so as not to advantage or disadvantage any providers.

3. Currently there are no clear significant influences of contextual factors on FfE PIs. However, if subsequent analyses of complete data identify any this may necessitate the: exclusion of PIs heavily influenced by context; adaptation of PIs to include contextual value added (CVA), or; provision of guidance to the interpretation of PIs most affected.

4. The LSC should display continued cognisance of issues concerning combination of PIs with narrow score distributions and regression to the mean, in the light of fuller analyses of data.

5. There are specific issues which relate to individual PIs. However the LSC is aware of them and has/is undertaking research and consultation exercises to try and address these. The results of these research and consultation exercises should be monitored against the potential risks to individual PIs identified in Section 7.2.3.

9. Following the launch of Version 1 of Framework for Excellence in June 2008, the next phase of activity, and a key challenge for the LSC and partners, is to ensure that the data and information gathered and presented under the FfE banner, is made available in a format that is accessible and appropriate to the needs of identified user groups. The LSC are currently addressing the ‘outward-facing issues’ of FfE data through the development of the Communications and Marketing Strategy for FfE. An integral part of this will be ensuring that FfE is used in the most fruitful way in order to drive up quality and standards, inform learner and employer choices, and serve as a self-improvement tool across the provider population.
1 INTRODUCTION AND APPROACH

1.1 Introduction and background

This is the final report of the Independent Review of Framework for Excellence (FfE) and details activity undertaken and the emerging findings. The Review began in May 2008, concluding in November 2008. The remit of the Review was based on the provider population included in the pilot activity and subsequent development, this did not include, for example, Personal and Community Development learning (PCDL) or school sixth forms.

GHK Consulting Ltd was appointed by the LSC National Office to undertake an Independent Review of FfE, with the intention of testing the robustness and fitness for purpose of FfE and its supporting methodology. More specifically, a ‘fresh pair of eyes’ were required to look into the extent to which the chosen performance dimensions and supporting indicators are:

- Appropriate for, and consistent with, the stated aims of Framework For Excellence;
- Consistent with the core design principles for FfE;
- Likely to give accurate, robust and discriminatory provider level scores that are stable over time.

The study was designed to be a rigorous and robust independent review of FfE.

The Review was informed initially by a series of key research questions, which can be found detailed in Annex A. As the Review progressed, the questions became condensed into three main areas for consideration:

1. Is FfE measuring the right areas?
2. Is FfE measuring these areas in the right way?
3. Are the data and results being interpreted in the right way?

Whilst all three questions have shaped study activity, the focus has predominantly been on questions 2 and 3. There is substantial evidence, including responses from the stakeholders consulted that the considerable amount of consultation, piloting, discussion and debate that has been invested in developing FfE thus far has ensured that the areas covered are the most appropriate ones, thereby responding to Question 1. Therefore focus on this area would have meant a significant amount of unnecessary duplication.

1.2 Study approach

A multi-method approach was taken in order to deliver the study. The main activities and related outputs are described below:
• Stakeholder consultation – following the inception stage, an extensive series of stakeholder interviews were undertaken, with individuals from a range of organisations including: Learning and Skills Council (LSC), Department for Innovation, Universities and Skills (DIUS), Office for Standards in Education (Ofsted), Quality Improvement Agency (QIA), Association of Learning Providers (ALP), Association of Colleges (AoC), National Audit Office (NAO), and the Audit Commission, as well as contractors involved in supporting various aspects of Framework development including IFF, KPMG, RCU, and York Consulting. A total of 21 interviews were completed. The information gathered through the interviews provided valuable feedback on the coverage, effectiveness and reliability of FfE from the perspectives of key partners, potential users, and wider sector representatives.

• Policy and document review – alongside the stakeholder consultation, GHK undertook a review of key documents relating to the development, and current shaping of FfE, as well as considering the policy context within which FfE operates. Relevant FfE documentation, reporting, consultation and strategic papers were included, as were materials relating to specific PIs such as the background to Training Quality Standard (TQS), or Ofsted inspection and reporting guidance. Materials relating to discussions at the FfE Technical Working Group and Policy Committee were also reviewed.

• Review of evidence on learner and employer choice – the study incorporated a review of existing evidence on what informs learner and employer choice, in order to support ongoing FfE development and to ensure that as far as possible, FfE data directly supports the learning and training decisions made by users such as employers and individual learners. One of the key functions of FfE is that it will “provide information to learners and employers to help them make decisions1”, this is clearly stated in the Principles of Design, Development and Implementation (outlined in Annex B).

• Comparative review of Performance Management/Measurement Frameworks – GHK also carried out a review of similar Performance Management / Measurement Systems, and frameworks designed to support quality improvement. The review examined approaches to specific operational functions such as gauging client satisfaction, gathering and using destination data, and more generally the procedures underpinning individual scoring and combining results to provide an overall score. The frameworks reviewed included approaches used by: the Audit Commission, Ofsted, the Healthcare Commission, the Commission for Social Care Inspection, Her Majesty’s Inspectorate of Constabulary, and the Quality Assurance Agency (QAA) for Higher Education. Documentation and research around inspection and assessment frameworks commissioned by the Department of Communities and Local Government (DCLG), the Higher Education Funding Council for England (HEFCE), the Department for Culture, Media and Sport (DCMS) and comparisons of financial dimensions commissioned by the LSC were also included. Whilst it is clear that the frameworks developed in different areas have specific purposes and operate within different contexts to FfE, there are still useful conclusions to be drawn from the comparison of systems.

---

1 Raising Standards and Informing Choice, LSC, March 2007
• Quantitative review of FfE – this involved a more detailed, quantitative based review of FfE as a whole, with a focus on individual PIs and general themes such as contextualisation, composite measures, combination rules and de minimis criteria. This review forms a crucial part of the study, addressing the question of how far the methodology supporting FfE is robust, defensible and reliable.

All of these activities have contributed to the Review, and informed the reporting and conclusions set out here. The remainder of the report is structured as follows:

• Section 2 provides an overview of the FfE, detailing user groups and their requirements, and goes on to highlight some of the issues raised during the comparative review, and stakeholder consultation.

• Section 3 introduces and reviews some of the cross-cutting themes which relate to FfE and performance frameworks more generally.

• Section 4 provides a detailed review of the key elements comprising the Responsiveness dimension of the framework.

• Section 5 presents a review of the Effectiveness dimension and the composite elements.

• Section 6 details and reviews the core components of the Financial dimension.

• Section 7 presents a summary and sets out key conclusions and recommendations.

• Annex A presents the key questions underpinning the Independent Review.

• Annex B presents the underpinning FfE principles of design, development and implementation.
2 OVERVIEW OF FfE

2.1 Background

FfE is a quantitative performance and quality measurement tool, developed to provide an independent, robust, and clear assessment of provider performance across the FE sector. The performance data supporting FfE covers a range of constituent parts, organised according to individual Performance Indicators (PIs), Key Performance Areas (KPAs), and Performance Dimensions, all combining to produce an Overall Performance Rating. The figure below illustrates the structure of FfE:

![Diagram of FfE structure]

Source: LSC FfE Website: [http://ffe.lsc.gov.uk/ffe/structure/](http://ffe.lsc.gov.uk/ffe/structure/)

Each stage of the development of FfE has thus far been based on a range of consultative and piloting exercises, more information relating to the activities informing the development of FfE can be found at: [http://ffe.lsc.gov.uk/ffe/](http://ffe.lsc.gov.uk/ffe/)
2.2 Current coverage – user groups and their requirements

The stakeholder consultation, Comparative Review of performance management systems, and the review of what informs learner and employer choice (see Section 1.2) formed the basis on which this Independent Review could make judgements regarding the current and potential strengths and weaknesses of FfE in its present format. FfE in its current iteration has a number of intended user groups including employers, individual learners, providers, LSC and other agencies, and IAG providers, and therefore is expected to perform a range of different functions, including:

- Acting as a source of information and quality indicators for learners and employers to inform choice;
- Providing an internal performance management tool for providers to inform and support self-improvement activity;
- Offering a facility for LSC/DIUS/Ofsted (and those organisations in future who will manage and commission provision) to monitor provider quality; and
- Presenting an overall sectoral picture to indicate areas for improvement and strengths generally across the provider population.

These functions are all seen as vital in supporting quality improvement across the sector. However, concerns were raised by stakeholders that the manifold functions of FfE could also present challenges, if the FfE was trying to be “all things to all people”. In addition, it was felt that there is a need to recognise the specific requirements of each user group in order to address them successfully. These are points to consider in the ongoing development of the Framework.

2.3 Overview of FfE – comparative review

A review of performance management and quality measurement frameworks was undertaken. This examined the experience of other sectors in developing performance frameworks, especially their approaches to assessing particular dimensions and specifying PIs and weighting these to arrive at overall judgements. The purpose of the review was to profile the development experiences of other frameworks, to inform the development of FfE and to underpin the ‘health check’ of FfE to see how it compares to established systems of a similar nature and purpose.

The main aspects of frameworks under review were:

- Balanced scorecard style service assessments: dimensions, weighting, ratings;
- Consumer and user surveys measuring responsiveness within performance assessments: how collected, how robust, how used;
- Judgements on particular areas of performance: use of PIs and inspection assessments, means of differentiation, meeting minimum standards.

These activities were supported by a review of literature search focusing on the web based publications of the Audit Commission, Ofsted, the Healthcare Commission, the Commission for Social Care Inspection, Her Majesty’s Inspectorate of Constabulary, and the QAA. This was supplemented by searches for past documentation and
research around inspection and assessment frameworks commissioned by the DCLG, HEFCE, the DCMS and others. A comparison report of financial dimensions commissioned by the LSC was also reviewed.

2.3.1 General emerging findings

In terms of drawing overall conclusions, the comparative review found that generally:

- The introduction of performance frameworks and ratings performance generates improvements in the outcomes being achieved (evident in schools and health service providers, and also in some though not all local authority services);

- Inspected and assessed bodies highlight similar benefits: self-assessment and scrutiny activities are encouraged, as is prioritising important areas of under-performance. Both of these tend to lead to much improved management information within the given organisation;

- Assessments of service which identify poor levels of performance tend to generate local public reactions, and can be an effective motivation for improvement. In response to poor assessment results, service users expect to see improvements ‘on the ground’ if they are to place value on the ratings;

- Ratings have generally improved following the establishment of performance assessment frameworks, however the benchmarks for performance have also been raised in some cases (evident in local authority service assessments, care providers, and in the core standards for health providers);

- Organisations are more likely to take action to improve in response to a process which clearly highlights causes for concern. Moreover, action responding to assessments is more likely if there are processes in place to ensure that follow-up action is being taken (for example some research identifies ‘coasting’ in some schools and local authorities who are in the middle ratings).

This suggests that: scoring approaches should in theory expose unsatisfactory performance, improvements in effectiveness and responsiveness should be expected before a provider’s ratings can improve, and future assessment of providers should encompass follow up action required to continue improvement.

2.3.2 FfE Performance Dimensions

In comparison with other identified frameworks, FfE rates particularly well in terms of the dimensions which constitute the scoring system. None of the other service assessment frameworks included in the review appear to be so clearly built on an equally balanced scoring of different dimensions (users’ views, inspectors’ views on quality of delivery and financial performance and user outcomes).

This lends a high degree of simplicity and transparency to FfE, especially when compared to other systems such as the assessment of local authority service areas and health services. Moreover, FfE appears to allocate a greater overall weighting to users’ views than, for example, the local authority services assessments.
Compared with the Audit Commission's guidance to the performance framework for local authorities, the component PIs and weightings of FfE used to arrive at scores and ratings are equally as clear. The system enables providers to identify where and how they need to improve to increase their resulting rating. Local authorities have become very adept at identifying areas for improvement and the targets to achieve in order to increase a specific rating, whether in terms of processes or service performance. FfE should equip providers with the same information and support them to improve where necessary.

2.3.3 FfE Responsiveness: user surveys

A key area for review was the incorporation of user surveys, within the framework of FfE and in comparison with other systems examined. The exercise found that user surveys have been used in the performance assessment of universities, registered social landlords (RSLs) and local authorities. In the case of local authorities, five examples were identified: a public survey (now known as the place survey) and surveys of library users, benefit claimants, planning applicants, and housing tenants. Key features common to these were identified:

- The practice of setting minimum numbers of responses for each service provider to ensure they are statistically robust, and allow for comparisons between providers over time. This minimum number can require some providers to undertake a census and others a sample;
- Prescribing the method for administering the survey in significant levels of detail, including the method of sampling and identifying dates for completion;
- Weighting responses from each provider/organisation to ensure they are representative of the population;
- Allowing respondent organisations to select additional questions from a ‘bank’ of pre-designed questions so that providers can incorporate individual requirements into the survey and avoid having to undertake alternative surveys;
- Setting clear benchmarks to distinguish levels of performance based on distribution of responses, so that for some services small differences in percentage points can translate into significant differences in overall scoring.

Whilst there are concerns which have been raised in relation to the reliability, cost and time required to administer learner and employer surveys within FfE, providing they are robust these surveys can be a valuable tool to indicate responsiveness. Results can also be used by providers themselves in their service planning to ensure they meet user needs.

2.3.4 Financial performance

The review found that the Audit Commission has undertaken a process of refinement of its assessment of financial health to cover a range of organisations in a similar way: local authorities, police forces, fire and rescue services, and health service providers. The approach is largely based on auditors assessing key aspects of financial performance against a template as part of their annual audit, rather than by largely assessing and weighting a set of financial indicators.
The approach adopted by the Audit Commission appears to be duly balanced, and the assessment encompasses both financial reporting and value for money judgements which look to be broader than the FfE approach. For example, the local authority financial assessment includes judgements on the achievement of economies in line with Treasury expectations. Much of this requires documentary assessment which is used by the Audit Commission and Ofsted to review ratings of service assessments in between inspections. How far this approach would fit within the FfE framework would require further investigation.

It is worth noting however, that the provider is given the opportunity to feed back contextual information to the LSC within the context of FfE, and the LSC has the facility to make decisions around grade moderation depending on specific circumstances e.g. large-scale investment in learning facilities which will ultimately benefit the learner but may affect the financial situation of a provider.

2.3.5 **Including different providers**

Most performance frameworks are devised to include a range of organisations, accommodating some significant differences in terms of size, user characteristics and range of functions. This could be said for care and health service providers as well as local authorities. A common assessment framework approach is maintained through:

- Employing a different balanced scorecard but similar ratings and assessment approaches: this approach is used to distinguish types of organisation within a category e.g. district councils from unitary and upper tier councils. Because such councils perform different ranges of functions this produces a more valid comparison between service providers. Within the FfE context, this may be appropriate for more specialist providers such as those only delivering Personal and Community Development learning (PCDL), for example, part of a local authority) or small work based learning providers;

- Taking out some elements of the calculation to arrive at a score where the element does not apply: this is used in the local authority assessment where counties and London boroughs do not perform the same functions. It is important in these circumstances only to do this where the calculation of a rating for a dimension will still be based on several indicators. This does not seem to be the case in the scoring of some types of provider in the Financial dimension of the FfE.

2.3.6 **Summary**

This section has identified the key points arising from the comparative framework review. Briefly, these comprise:

- The balanced scorecard approach of FfE should deliver benefits to providers which are similar to those in other sectors. It should also drive up performance, and ensure that improvements in responsiveness and effectiveness are required to increase overall ratings;

- Post-inspection action planning and ensuring ongoing improvement in delivery are instrumental in supporting continuous improvement;
The FfE is relatively clear and transparent and should be an effective tool for providers to self-improve, although it should be recognised that generally, larger providers will be more adept at using and interpreting performance data;

The FfE gives a higher weighting to responsiveness. The learner survey should be robust providing it has the key ingredients listed above and is administered in a transparent manner, according to LSC prescribed conditions;

Financial performance can be more broadly defined and have a lower reliance on PIs which could be disputed as measures of financial performance if additional assessments are carried out by auditors. The LSC has the facility to moderate financial scoring in certain circumstances for a provider;

It is possible to include different types of provider within the same performance framework provided only small parts of a particular dimension exclude them where appropriate. It should be possible equally to differentiate the balanced scorecard for some types of provider but retain the same approach and overall rating format.

2.4 Overview of FfE – stakeholder perspectives

The stakeholder consultation yielded some very positive comments in relation to FfE, underlining the desire for a single quality and performance management system such as this. As far as stakeholders were concerned, the majority felt that the intentions and overall purposes of FfE were worthwhile and that the principles underpinning the Framework were sound. Bringing together quantitative information relating to provider performance, and provider quality, within a user-friendly, clear and robust framework for reference was seen as a desirable development.

To a certain extent the stakeholder consultation provided a good level of reassurance that the FfE is fit for purposes to achieve the stated aims and objectives in a robust way. However, alongside the positive responses to FfE, some concerns were voiced with regard to particular aspects of the Framework, and some specific questions raised. These are outlined below, along with relevant findings from the wider stakeholder discussions, comparative framework review, and review of what informs learner and employer choice.

2.5 Issues arising from stakeholder consultation

2.5.1 Issue 1 - Reliability of indicators and measures

A crucial factor determining the overall value, and the credibility of FfE is the reliability and statistical robustness of the data presented. For Framework for Excellence data to be effective across the uses described in section 2.2, all user groups must have confidence in the information they are accessing. Generally, stakeholders were satisfied with the reliability of FfE data. However, concerns were expressed, for example about the use of a sample-based approach to the learner and employer satisfaction PIs. Survey use within FfE is presented and discussed under Section 3 as a cross-cutting issue.

Responding to the issue

The fact that FfE scoring is based on quantitative data presented according to the individual PIs provides a good degree of robustness. The use of sample based
surveys represents the main area of concern amongst stakeholders as regards reliability of data.

The web-based survey approach goes some way to addressing these concerns. Using the internet as the platform for responses means that the surveys are not tied to a set sample of respondents, and that all of the population could potentially respond to the survey. However, response rates may be lower.

The LSC has recognised the importance of robustness and reliability, and has invested considerable time and resource into exploring the key strengths and weaknesses of chosen approaches. Key activities include the ongoing input of the FfE Technical Working Group, widespread consultation, targeted research addressing key areas such as contextualisation, combination rules, etc. and also the piloting exercises undertaken to inform further development. All of these have contributed to the LSC recognising the importance of having valid and reliable indicators and measures.

2.5.2 Issue 2 - The value of FfE data to users

Ensuring the value of FfE data to users, especially given the range of users and their respective requirements, is a key challenge for the LSC and partners. The suitability of FfE data to core users such as employers and learners in order to help them make informed decisions about their training and learning options is paramount. Reaching a point where the data is available in a suitable format, at an appropriate level (geographical, by course/Department etc.) and will be used in the most fruitful way is a key objective in the ongoing development of FfE. The LSC are undertaking work geared towards ensuring the data collated through FfE is presented and used in the most appropriate way according to the needs of identified users.

Responding to the issue

Learners and employers - The Review of What Informs Learner and Employer Choice found that learners and employers place emphasis on a number of areas when making their training decisions. These include success rates, learning outcomes, and destinations for their chosen course or qualification, quality of the provision (e.g. quality of facilities), and information such as costs, access criteria, and mode and timetable of delivery.

Table 1 below offers an indicative overview of FfE measures, and cross-references these with some of the learner and employer information needs each will address.
## Table 1 Overview of FfE measures and learner and employer information needs

<table>
<thead>
<tr>
<th>Dimension</th>
<th>KPA</th>
<th>PI</th>
<th>Interests addressed</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Effectiveness</strong></td>
<td>Quality of outcomes</td>
<td>Success rates</td>
<td>Learner interest is high in achievement rates for a course or programme. Equally, employer interest is high in achievement rates for a specific programme and/or course. Both groups will also be interested in the success rates of a provider overall.</td>
</tr>
<tr>
<td></td>
<td>Quality of provision</td>
<td>Inspection grade</td>
<td>Ofsted reporting offers more qualitative judgements about provision than FfE data alone. Ofsted judgements provide an extra dimension to learner and employer choices. Incorporating Ofsted rating into FfE means both sets of information are accessed at once.</td>
</tr>
<tr>
<td><strong>Responsiveness</strong></td>
<td>Responsiveness to learners</td>
<td>Learner views survey</td>
<td>Learners attach importance to other learners’ experience of a provider (either through word of mouth or through testimonials, etc.) The survey directly addresses this and offers a view from previous learners. There are questions around sampling and administration, but the value of direct consultation with learners outweigh these.</td>
</tr>
<tr>
<td></td>
<td>Learner destinations</td>
<td></td>
<td>Gives learners an idea of the uses of a qualification, and offers background on factors such as increased employability, leading on to other training, etc. Can help to identify progression routes for learners.</td>
</tr>
<tr>
<td></td>
<td>Employer satisfaction survey</td>
<td></td>
<td>Employers need to know that a provider can meet their requirements. Knowing that a provider has been highly rated by other providers can help to make decisions. Again, there are questions around sampling and administration, but the value of direct consultation with employers outweigh these.</td>
</tr>
<tr>
<td></td>
<td>Amount of training</td>
<td></td>
<td>The value of this data to inform learning decisions of employers and learners is not high. However, the value for the LSC and other commissioning bodies is clear.</td>
</tr>
<tr>
<td></td>
<td>Achievement of the Training Quality Standard for Employer Responsiveness</td>
<td></td>
<td>For employers, knowing that a provider has been awarded the TQS is a clear indication of quality and responsiveness. The use of this external judgement within FfE means that an extra dimension of reassurance can be provided.</td>
</tr>
<tr>
<td><strong>Finance</strong></td>
<td>Financial health</td>
<td></td>
<td>The financial indicators are of interest, however evidence suggests that neither employers nor learners place high levels importance on receiving information on finance related data to make their training decisions.</td>
</tr>
<tr>
<td></td>
<td>Financial control</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Use of resources</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Therefore for informing the decisions of employers and individuals, FfE can address some of the areas of interest identified by previous research (e.g. success rates), quality of provision (Ofsted Inspection Grade), user satisfaction (learner and employer surveys). However, there are some information needs that will not be addressed by FfE data: enrolment/access criteria, mode and timetable for delivery, and course costs.

Also emerging from the review of evidence, and highlighted by some stakeholders, was the fact that both learners and employers’ decisions are often influenced by other factors such as: their prior knowledge or experience of a certain provider, word of mouth recommendations by peers, colleagues, or other businesses, the location of a provider and the ease of access, and sectoral strengths in their area of interest. These factors are outside of the scope of influence of FfE.

Providers – FfE data is also intended to support providers in their quality improvement activity, and to underpin the move towards self-regulation for the sector. Stakeholders felt that FfE data would be useful for providers to identify their strengths and weaknesses but also emphasised the need for support and advice from agencies in achieving the improvements, e.g. action planning and ongoing guidance. The Ofsted post-inspection action planning process was cited as a good example of this.

Research is currently being undertaken on behalf of the LSC aimed at scoping and identifying the information needs of the potential user groups for FfE. This will directly inform the development of the outward-facing elements of FfE, addressing issues such as drill-down facility for data, format and presentation of information and levels of use for different types of information.

### 2.5.3 Issue 3 - The coverage of FfE and component areas

Generally, the stakeholders consulted felt that FfE covered those areas necessary to provide a picture of the overall quality and performance of a provider. However, suggestions were made that other areas would enhance the coverage of FfE, such as the responsiveness of a provider to the community in which it delivers. This relates directly to the Comprehensive Area Assessments developed by the Audit Commission which adopts a holistic approach to area review, rather than focussing on a specific service or facility. The additional areas suggested below present some elements for consideration in the future. However during these early stages of development for the Framework, stakeholders believed that keeping the FfE as simple as possible will add to its effectiveness in the immediate future.

**Responding to the issue**

As highlighted in section 2.3.2, the coverage of FfE compares favourably with other established performance measurement frameworks. The mix of outcomes and quality assessments in the FfE under the measures relating to effectiveness closely matches the assessment of the effectiveness of local authority services. In terms of discriminating performance, the Audit Commission’s experience generates a number of suggestions. On a general level, the Audit Commission found that the more complex the composites that are combined to produce overall performance area scores, the greater the likelihood of the provider manipulating performance or ‘gaming’. Also, they target resource investment in trying to manage the system to increase their scores and focus on particular areas of activity to increase their rating. FfE compares well with this
regard, as two of the three main areas of assessment are based on the experiences of users, and on outcomes, and therefore are less susceptible to manipulation.

Therefore, measuring the effectiveness of provision under FfE, i.e. the quality of outcomes and quality of provision, in some ways reflects approaches adopted elsewhere. However there are differences between the FfE and these other approaches, which is to be expected given the nature of the frameworks and the fact they have been developed and tailored to measure provision of particular services.

2.5.4 Issue 4 - Focus on measures rather than performance

Some stakeholders felt that the FfE could create a focus on measures rather than quality and performance more widely, and there was also a risk of PI driven perverse behaviour amongst providers. For some, this was compounded by the absence of wider ongoing support for providers in terms of action planning guidance to steer targeted improvement, facilitation of good practice examples, and supporting peer learning. However, it should be noted that this was not seen as a significant risk across the provider population, and that the communication and dissemination activity supporting FfE will help to address these concerns. Similarly future developments such as the planned benchmarking facility based on the ‘statistical neighbour’ approach will support providers in their improvement activities.

Responding to the issue

A clear point to make regarding this issue is that, providing the areas measured are those which contribute to provider quality, there should be little cause for concern. As outlined previously, the comparative review identified key related points emerging from other frameworks:

- Introducing performance frameworks and ratings generally leads to overall performance improvement;
- Self-assessment and scrutiny activity encourage a focus on addressing areas of under-performance, leading to improved management information;
- Assessments highlighting examples tend to generate local public reactions and are a strong motivation for improvement;
- Ratings have generally improved across the frameworks reviewed, even in examples where benchmarks for performance have also been raised;
- Action to improve is more likely to be taken where the process highlights causes for concern and follow up checks are undertaken.

These points suggest that: scoring within frameworks should expose unsatisfactory performance; improvements in effectiveness and responsiveness should be necessary before a provider’s ratings can go up, and consideration should be given to providers taking action to improve, and supporting such an action planning process. More generally, it would be expected that, as confirmed by the review, providing the areas measured are the correct ones, then overall quality would be improved by default.
2.5.5 **Issue 5 - Burden of reporting**

Recognising that many providers already have established systems and processes for data collection and analysis to support self-improvement, e.g. surveying learner satisfaction, a key concern expressed was that FfE would signal an increase in the burden of reporting for providers.

**Responding to the issue**

As far as possible, the consultation phase has taken account of provider opinion on the proposals, and the practical implications of reporting required at each stage have been considered. Some data requirements will inevitably place demands on providers, however the LSC has produced guidance and has made support available to help providers to address this.

Emerging from the Pilot Evaluation Reporting were two key messages. Pilot providers responding to the consultation were generally “pleased to be involved in such a large-scale and fully consultative project”. However, with regard to the operational aspects of their involvement, the survey found that “The costs and burden of administering the Learner and Employer Views surveys were seen as fairly high. Some providers perceived questionnaire fatigue and would like the Framework surveys to be “piggybacked” onto their own local surveys” (p15). The LSC is seeking legal advice about the possibility of doing this.

2.6 **Section summary**

This section has provided an overview of FfE more generally, and of the comparative review of Performance Management Frameworks, the review of what informs learner and employer choice, and the stakeholder consultation exercises. Key points emerging include:

- FfE has been developed in order to respond to the needs of a number of specific user groups, and as such performs a range of functions. The data presented by FfE will be used for a variety of purposes, including informing learning choices, performance management, commissioning and procurement, and sectoral analysis.

- The comparative review found that FfE stands up well against other established performance management frameworks. Key points include:
  - the use of a clear balanced scorecard approach with transparent weightings across performance dimensions;
  - measuring responsiveness to users through direct consultation. Whilst not without its issues, the direct consultation approach is a valuable one in ensuring needs of users are met;
  - the FfE financial appraisal does not include auditor judgement, however the LSC has a facility to moderate scoring according to specific provider circumstance;

---

FfE includes different types of provider within the balanced scorecard approach, but retains continuity through the same approach and overall rating format.

The stakeholder consultation identified support for FfE across the board, and general backing for a single source of performance data on providers. A number of issues were also highlighted:

- Reliability of data, indicators and measures
- The potential value of FfE data to user groups
- The coverage of FfE and component areas
- A focus on measures rather than performance
- The burden of reporting

All of these issues have been considered and discussed by the partners developing FfE. In some cases, e.g. the reliability of data, especially using surveys to generate data, the issues are an inevitable part of the approach and whilst they should be addressed as far as possible, are unavoidable.
3 CROSS CUTTING ISSUES

3.1 Introduction
Based on the main issues identified by the stakeholder discussions, the comparative review of systems, and the review of what informs learner and employer choices, this section presents and discusses some of the main operational cross-cutting themes which affect the development and implementation of FfE. Essentially these comprise:

- Contextualisation
- Use of surveys
- Use of composite measures
- Combination rules
- *De minimis* criteria and exemptions
- Balance of weighting

3.2 Contextualisation
Contextualisation issues represent a common complication for most performance management frameworks in which providers are compared against national benchmarks rather than uniquely against only those providers that share the same context e.g. socio economic conditions of location.

General research to investigate the methodological challenges (including contextualisation) faced in developing performance assessment frameworks, such as in healthcare, local government and corporate performance assessment, have been conducted by the Centre for Health Economics\(^3\) and by the Audit Commission.

The Audit Commission examined this issue at the outset of corporate performance assessment and considered that because the funding system for local authorities took account of needs in most services, areas with more needy populations, (for example, higher demand and higher delivery costs) were better funded to meet these needs. The Commission has attempted to alleviate the problem by:

- Avoided PIs which are heavily influenced by context (these are generally input PIs which have generally not featured in the assessment);
- Arrived at scores for some PIs based on the distribution of the PIs for types of local authority (metropolitan, London borough, unitary and county) so having different lower and upper thresholds for each type;

---

\(^3\) *Public Services: Are Composite Measures a Robust Reflection of Performance in the Public Sector* (2006). Centre for Health Economics, University of York, York, UK.
- Guided inspection judgements in some cases to take account of the population served and the difficulties this gives rise to.

There are specific situations where the Audit Commission adjusts PIs for 'context'. This is where there is a strong statistical correlation between performance and deprivation.

The development of FfE on the other hand, has incorporated in-depth work on behalf of the LSC exploring the impact of contextual factors and offering judgements on how far these should be taken into account when deciding scores. This research identifies contextualisation as an issue of concern when "results can be generated for a provider but these are skewed in a predictable and measurable way by an external influence unrelated to performance".4

While preliminary findings of the research suggest that there are no clear significant influences of any socioeconomic contextual factors (such as ethnicity, local educational attainment, population density) or provider type (such as, degree of specialism) on the ratings achieved for any of the performance indicators in the FfE, the reassurance this might otherwise provide is limited by the fact that these findings are based on limited data from the pilot. (Level of course did affect learner satisfaction ratings, with those on higher level courses tending to give lower satisfaction ratings). Furthermore, specific findings suggest that ratings on each of the performance indicators are partially and weakly significantly related to contextual factors such as ethnicity, deprivation level and area prior attainment and, as will be discussed with individual PIs, there is at least an a priori basis for anticipating that more complete data may exhibit more marked effects.

If, within a year, analysis of the large-scale more complete data available from the FfE research does not dispel remaining concerns, it is suggested that one or more of the following actions may mitigate the effects of contextualisation:

- Exclusion of PIs which are heavily influenced by context;
- Include a control group;
- Adapt the PIs to include contextual value added (CVA);
- Provide guidance to the interpretation of the PIs most affected.

3.3 Use of surveys

3.3.1 Survey implementation

Learner surveys

As far as other frameworks are concerned, user surveys have been employed in a number of contexts, including in the performance assessment of universities, RSLs and local authorities. For local authorities there are five: a public survey (now known as the place survey) and surveys of library users, benefit claimants, planning

applicants, and housing tenants. In general, development of high quality surveys are underpinned by several key features:

- Setting minimum numbers of responses for each service provider so that all PIs based on surveys are statistically robust and can be compared between providers and over time; this would require some providers to do a census and others a sample;

- Prescribing the methodology in detail including the sampling method and dates for completion;

- Weighting responses from each provider to ensure they are representative of the population;

- Allowing additional questions from a ‘bank’ so that providers can piggy back on the surveys and avoid any alternative surveys;

- Setting benchmarks to distinguish levels of performance. For some services small differences in percentage points can lead to significant differences in the scoring.

Within the context of FfE, while we understand there are implications about the cost and time required to administer such surveys, they need to be robust to have value so that they can be used by the providers themselves in their service planning. A more detailed review of the surveys used in the FfE is provided in sections 4.1.1 and 4.2.1. However, it is worth pointing out at this stage that their development appears to have been conducted with full cognisance of the features bulleted above.

**Employer surveys**

The scoping exercise found that there does not appear to be a similar survey undertaken in other performance frameworks. By and large all surveys in other performance frameworks are only of direct users; for example, library users, benefit claimants, patients, third year students. The exceptions are:

- Seeking the views of current parents about the quality of a school; though this is not by a systematic survey method and is only used to inform and guide inspection;

- Seeking the views of the public in general about the quality of services which are not universal and may not have been used recently (local authority residents’ survey).

The lesson appears to be that the survey must be of recent users of a service. In FfE, all employers surveyed are direct beneficiaries through the LSC’s Employer Responsiveness funding model (many of which are funded through Train to Gain). Employer surveys ought to be amenable to the same requirements as learner surveys to be robust – minimum achieved response, weighting, standard method, defined population. This suggests that the survey should generally be a census if the numbers are small.
3.3.2 **Survey Representativeness:**

The question of representativeness essentially relates to the recurring theme of confidence intervals. The most common use of confidence intervals is to give an estimate of the extent to which the score for a measure, based on the sample, is representative of the whole population of respondents. The usual confidence interval employed is that of 95 per cent. Put simply for this PI, if we were to repeatedly (randomly) sample 100 learners from a provider with, for example, a total population of 500 learners, what would be the range of mean satisfaction scores we would obtain 95 times out of 100? Clearly, the higher the proportion of learners in the sample (out of the total population of learners), the smaller the range (interval) of mean scores is likely to be because each repeated sample would contain many of the same respondents (who, we can assume, would respond in the same way each time).

In the future, confidence intervals (and representativeness) may become a less significant issue given that the LSC will be encouraging providers to extend the 2009/10 learner views survey to all learners. The fact that future surveys will be all-inclusive highlights another concern: the tension between obtaining high (or complete) sample sizes and increasing bureaucracy for providers, many of whom are already conducting their own learner surveys. One possibility to ameliorate this problem is to combine the two via the LSC website.

For the present, the proportion of learner respondents within different providers will have an impact on both the statistical robustness of the measure, especially for small providers, and also the extent to which specific views are over or under-represented (through non-response bias). Representing the views of learners with learning difficulties is also particularly problematic, given the discretion providers have in including (or excluding) their responses.

The issue of confidence issues has been further, and somewhat unnecessarily complicated by the concept of super-populations. In the same way that we can query the extent to which a random sample of learners is representative of the whole population of learners at a provider, we could also query the extent to which the population of learners at a provider are representative of all possible learners (the hypothetical super-population). The concern here is that we may well obtain different satisfaction scores for an individual provider, year-on-year, but we may not be able to infer that the level of support or provision has changed year-on-year because it may be that it is the propensity of the learner population to be satisfied that has changed instead. The only way to estimate this kind of “non-sampling error”, and hence estimate how much we would expect the learner satisfaction scores to vary year-on-year due to changes in the population of learners (and not changes to underlying levels of support and provision), is to analyse survey data over a series of years.

In practice, however, there is good reason to question how much (or little) one should reasonably expect the population of learners to differ year-on-year at an individual provider in terms of their propensity to be satisfied. We might expect contextual factors (such as local socio-economic and labour market changes) and changes in provider admission policies (i.e. introduction of selective admission) to impact on the type of learners at a provider year-on-year. We have dealt already with the issue of contextual factors (see section 3.2). However, the issue of changes to admission policy may be dealt with relatively straightforwardly by simply flagging up those providers for whom such changes have been implemented.
In the absence of acute characteristic changes in socio-economic and labour market factors or provider admission policies, there is little reason to suspect that the population of learners at an individual provider differs significantly from one year to the next in terms of the hypothetical super-population from which they could be derived. Given the intention that the FfE be used more to provide yearly ‘snapshots’ rather than longer term trends, the whole issue of super-populations becomes an even more redundant one.

3.4 Composite measures

Combinations of PIs to make up composite scores for performance dimensions are common to all the frameworks reviewed. It is clear that the weightings used and the way scores are arrived at from PIs and other information can have a significant effect on the rating. As will be discussed in section 6.4 on general issues concerning the financial dimension of the FfE KPAs, composite indicators are aggregated measures comprising a number of sub-measurements.

While they have the advantage of simply conveying complex information, they also raise a number of disadvantages, most of which are discussed in detail by the Centre for Health Economics. It is not the intention to revisit those detailed and lengthy investigations here, but simply to relate some of the specific challenges they identified to the Framework for Excellence and discuss possible ways in which the Framework’s PI methodologies might be adapted to mitigate these.

Some of the methodological, interpretive and behavioural challenges of composite measures, as they relate to the FfE include:

- Aggregation of individual performance measures may disguise serious failings in some parts of the system. More importantly, composite KPAs make it more difficult to identify the strengths and weaknesses of individual providers, which impacts on their potential for remedial action.

- As we will see, some of the PIs are more statistically robust and reliable, more appropriate to what they intend to measure, and less context sensitive than others. Combining PIs of variable quality essentially reduces the composite measure to the lowest common denominator.

- Conversely, simply excluding the poorer PIs from composite measures can distort the true underlying performance of some providers.

- Composite measures are highly sensitive to the weightings attached to individual PIs, which in turn may not be formulated with any clear rationale or evidence base. In the case of some of the more empirical PIs, many of the weightings are self-explanatory and justified (as is the case in the calculation of the Qualification Success Rates). In others, however, the weightings may be more arbitrary.

- Composite measures are also highly sensitive to combination rules (see section 3.5), which in the FfE, are themselves a potentially contentious issue.

Various strategies can be employed to minimise the risks incumbent on composite measures, many of which would be usefully employed in the FfE:
• Rules on combination should be subject to careful development and high levels of scrutiny as fairly subtle changes in these may have a profound impact on the ratings, and rankings, of individual providers. Key to this is the need to investigate, using as much genuine previously collected data as possible, the actual impact that the various possible combination rules might have.

• Relatively few PIs should be based on sampling; where sampling is used the sample size needs to be robust and the method applied needs to be consistent;

• Weighting systems should be based on clear rationales (as is the case for the Qualification Success PI; see section 5.1.1) rather than arbitrary rules. Furthermore, weightings should not vary from year to year. This issue is relatively unimportant in the FfE as the PIs are given equal weighting, although it might be worth exploring whether in fact they should be given equal weighting if some can be deemed more influential than others.

• Composite measures should be published with indications of the levels of uncertainty of the underlying PIs, and consequently the compounded uncertainty of the composite.

• As far as is possible, it is better to first summate (after weighting) as many standardised raw scores as possible, before converting to a simple rating system, than it is to convert individual raw measurements into rating systems and then combining them. As a simple example, current GCEs are comprised of six units, each assessed out of varying numbers of raw marks. While raw marks for each unit may be reported as grades for each unit (so that stakeholders can see how candidates perform in specific areas), the overall A-level grade is based on summation of uniform (standardised) marks first, and then conversion to a grade, rather than combination of individual grades.

The combination rules developed in support of FfE reflect the importance of these issues.

3.5 Combination rules

Following on from the discussion of composite measures above, one of the key issues with regard to the FfE is how individual PIs are combined to produce grades for key performance areas, which are in turn combined to provide grades for the dimensions, which are finally combined to provide a grade for the overall performance rating (OPR). Considerable time and resource has been invested in exploring the debates and issues surrounding combining PIs into composites.

New combination rules for the FfE have now been agreed. However, it is worth discussing briefly what the key concerns are. Of particular interest, is which of two main methods of combining PI ratings, is used:

• combining performance scores or,

• combining weighted grades.5

A paper to the Policy Committee issued in April 2008 recommended that combining point scores (the former method) was preferable to combining grades (the latter method) as the former allows considerably greater discrimination in OPR scores. This is basically the reasoning behind the aggregation of uniform marks from A-level units to derive an overall A-level grade, as described in the section on composite measures above.

One of the other concerns relates to the tendency of the combined scores, being averages of other scores, tending to cluster more narrowly around a central point. Again, while this issue has been discussed in relation to the Learner Views Survey and the Employer Satisfaction Survey, it is actually further compounded when scores are combined. Essentially, the issue is one of “regression to the mean”: ideally, we would expect a set of scores from each PI to be distributed a bit like a bell-shaped curve, with the majority of providers having scores around the mid-point (the mean) through to a few providers having extremely high or extremely low scores. This would be our (“normal”) distribution of scores. However, if we were to combine a set of scores for each provider, we would find that combined scores would now be more narrowly distributed (clustered more closely to the mean combined score) than is the case for individual PIs. This is a common statistical phenomenon, however, as a general rule, the greater the number of individual scores that are combined, the greater the clustering (regression) towards the mean. This problem is even more compounded if any of the individual scores are themselves narrowly distributed, as is the case for the Learner Views Survey and the Employer Satisfaction Survey.

The two main ways of dealing with this problem are to either adjust the cut-scores (assessment criteria) to align them with the narrowing distribution of OPR scores, or to adjust the scores to introduce more variability in the extremes (high and low scores). On the basis of the results of the (unrepresentative) sample of all 100 pilot providers, it has been agreed that the latter method (adjusting the scores to introduce more variability), be adopted, subject to confirmation or variation of the parameters of the adjustment once the version 1 full data set is available in spring 2009.

Essentially, the adjustment to scores to introduce more variability (i.e. increase the proportions of providers in the OPR Grade 1 ‘outstanding’) takes the form of allocating Additional Performance Points (APPs) at OPR level based on the number of KPA scores above 60 (and a minimum PI score being above a threshold of 40).

Present analyses of the available data and data simulations suggest that this method does indeed improve the distribution of OPR grades somewhat and has an ameliorative effect on problem of regression to the mean and narrow score distribution.

### 3.6 *De minimis* criteria and exemptions

Most performance frameworks are devised to cover providers with some significant differences in terms of size, user characteristics and range of functions. This could be said for care and health service providers as well as local authorities. The FfE similarly covers a range of providers, and with further roll out this will be even more of a diverse set of provider types, sizes, and characteristics. Some issues arose through the development process, and were also put forward by stakeholders, that FfE does not apply easily across all types of providers. These concerns have been taken on board by the LSC, and reflected in factors such as the exemption criteria.
De minimis criteria essentially refer to these rules of exemption applied to providers from participation in individual PI evaluations. Most of these are based on excluding providers with low numbers of learners or those for whom the proportion of LSC funding is low.

One problem is that the FfE is viewed by some as being largely predicated on an FE College model; other training providers are quite different, and the level and quality of their data is not necessarily as robust. Therefore the information may be available for FE providers, but may not be as clear for work-based learning providers. It was suggested there is a need to undertake research into weightings and how, if at all, these can be affected by exemptions, whether these offer particular advantage or disadvantage to providers and in which cases. Weighting is crucial, as is ensuring that exemption levels do not affect the overall balance of results in terms of offering advantage or disadvantage to providers, and also being clear on exactly what value each one gives and what this means in terms of quality. An associated issue is that de minimis rules may encourage some training providers to reduce capacity below exemption levels and so avoid having to produce certain PIs. This relates back to the wider issue of driving perverse provider behaviours, discussed in more detail elsewhere in the report.

There is a more general issue concerned with how exemptions are handled under the rules of combination. There are three ways that exemptions can be treated:

- Attribute a score to the exempted value, based on representative performance by the provider elsewhere in the framework, this is essentially how, in the past, candidates who missed an A-level unit examination through illness or injury were treated;

- Ignore the exempted indicator, resulting in the remaining aspects of performance in that dimension effectively having greater weight in determining the dimension grade for those providers;

- Ignore the exempted indicator, but compensate for its loss by explicitly reducing the weight of the dimension accordingly in determination of the OPR.

The LSC’s own research suggests that because the first method would give nonsensical results, with for example, SFCs appearing to have excellent employer responsiveness when they are not serving local employers, and the third method deviating from the policy statement that each dimension should be of equal weight, the second method should be adopted.

The question of how far a generic system for measuring performance can be applied across the provider population remains. Clearly, there is a strong interplay between de minimis issues and combination rules and it should be highlighted that the new combination rules developed for the FfE appear to rectify all of these problems making the need for any further actions redundant.

3.7 Balance of weighting

Development of the FfE represents a departure from many other performance management/measurement frameworks or earlier models, in that it has a fairly broad assessment base, including Key Performance Areas of Responsiveness to Learners,
Responsiveness to Employers, Quality of Outcomes, Quality of Provision, as well as the Financial Key Performance Areas.

Consequently, compared with the past and many other frameworks, the Financial Dimension in the FfE is much diluted within the OPR. However, the FfE also has a broader set of users than many other frameworks and it could be convincingly argued that the inclusion of other dimensions appropriately reflects this breadth.

Nevertheless, the question remains that the FfE policy to give equal weighting to each of the three dimensions, should, perhaps, not be accepted without comment. While this enables providers to identify where and how they need to improve provision to increase their resulting rating, it also raises an issue that is, to some extent, a philosophical one: equal weighting would imply that each of the individual components are equally important in determining the OPR, and are therefore equally important in reflecting the FfE aim of providing an independent, robust, and clear assessment of provider performance across the FE sector. Given, however, that defining the relative importance of the different dimensions is perhaps a moot point, one could argue that giving them equal weighting provides a practical, and arguably, a parsimonious solution.

One consequence, however, is the extent to which this skews the resulting OPRs of those providers who do not, generally, have calculable PIs for all KPAs. Sixth form colleges, for example, are for the most part exempt from providing Employer Responsiveness ratings, which represent $\frac{1}{6}$ of the scores for the OPR. It is difficult to see how this can be rectified without upsetting the balance of scores for other providers and is perhaps, therefore, more an issue that those interpreting the data should be cognisant of, with appropriate interpretative guidance from the FfE.

3.8 Summary

The development of any performance management/measurement system or framework, and consequently any quantitative review of its quality and utility requires cognisance of issues that will impact across specific Key Performance Areas or Performance Indicators. Consequently, while the discussion of cross-cutting themes in this section of the review is a fairly detailed analysis of potential concerns, this does not imply that the issues discussed have been in any way neglected in the development of the FfE specifically – in many cases it is quite the contrary – or that any specific action, over and above cognisance of them, is required. It is therefore useful to summarise the key issues as they relate to the six (somewhat interrelated) cross-cutting themes and to indicate where, if any, we recommend any specific action to be taken in the FfE, or where the concern has been adequately addressed:

3.8.1 Balance of weighting:

- Issues:

  1. The FfE has a broad user base which is reflected in a broad measurement base in which dimensions are given equal weighting, which may or may not be an accurate or correct reflection of the relative importance of different dimensions on provider performance.
2. Providers for whom certain KPA data are not collected may be unfairly advantaged or disadvantaged because of the increased relative importance of the other dimensions.

- **Recommendations:**

  1. FfE should provide guidance in the interpretation of OPRs where providers are exempted from an entire KPA.

3.8.2 **Contextualisation:** impact of contextual factors (e.g. ethnicity, local educational attainment, population density etc.) on performance scores.

- **Issues:**

  1. Initial findings of analyses suggest there are no clear significant influences of contextual factors on FfE KPAs or PIs, this is based on limited data available from the pilot.

- **Recommendations:** depending on evidence from analyses of complete data, it may be necessary to implement one or more of the following actions:

  1. Exclusion of PIs heavily influenced by context;
  2. Adaptation of PIs to include contextual value added (CVA);
  3. Provision of guidance to the interpretation of PIs most affected.

3.8.3 **Use of surveys:**

- **Issues:**

  1. Minimum number of responses should be set for providers;
  2. Survey of smaller providers should be based on complete census rather than selective sampling;
  3. Method should be prescribed in detail, including method of sampling and dates for completion;
  4. Responses should be weighted from each provider to ensure they are representative of the population;
  5. Benchmarks should be set to distinguish levels of performance;
  6. Additional questions should be available from a ‘bank’ so that providers can ‘piggy back’ them on their own surveys.

- **Recommendations:** none as issues have been addressed, although in the case of issue 6 (‘piggy backing’ onto providers’ own surveys), the action has been deemed unworkable, at least in the near future.

3.8.4 **Use of composite measures:**

- **Issues:**
1. Aggregation of individual performance measures may disguise serious failings in some or part of the system;

2. Combining PIs of variable quality essentially reduces the composite measure to the lowest common denominator;

3. Excluding poorer PIs can distort performance scores;

4. Composite measures are highly sensitive to weightings, which should therefore be formulated with clear rationale;

5. Relatively few PIs should be based on sampling.

- **Recommendations:** no specific recommendations as the FfE already has comparatively few PIs and KPAs and has developed them with cognisance of the issues.

### 3.8.5 Combination rules:

- **Issues:**

  1. Combining point scores is preferable to combining weighted grades, as it allows greater discrimination in OPR grades.

  2. Regression to the mean, and combining of some PIs with very narrow distributions results in OPR ratings clustering towards central values (of 2 and 3).

- **Recommendations:** none as the FfE has adopted measures to deal with both issues, and intends to review these in the light of complete data sets.

### 3.8.6 De minimis criteria and exemptions:

- **Issues:**

  1. FfE is viewed by some as being largely predicated on an FE College model; other training providers are quite different, and the level and quality of their data is not necessarily as robust.

  2. Weighting is crucial, as is ensuring that exemption levels do not affect the overall balance of results in terms of offering advantage or disadvantage to providers.

  3. *De minimis* rules may encourage some training providers to reduce capacity below exemption levels and so avoid having to produce certain PIs. This relates back to the wider issue of driving perverse provider behaviours, discussed in more detail elsewhere in the report.

- **Recommendations:** none, as the new combination rules reduce the impact of all of these issues to negligible levels.
4 RESPONSIVENESS DIMENSION

4.1 Key Performance Area: Responsiveness to Learners

4.1.1 Performance Indicator: Learner Views Survey

The main purpose of this PI is to present qualitative information including "learners’ perceptions about the information, advice and guidance they received from their provider, the quality of teaching and learning ... [and] their learning experience" in terms of a quantitative rating measure. The usual methods of measuring this kind of qualitative information would include one-to-one structured or semi-structured interviews, focus groups, consultation events and questionnaires/surveys. Given the large number of learners in the sample, the survey method actually employed seems entirely appropriate. The key questions that arise are:

- Does the way in which the survey is administered advantage or disadvantage different providers?
- How well is the survey constructed as a measurement tool?

Survey administration

In terms of administering the survey, there are issues attached to each option of either the providers taking responsibility, or the LSC (through an independent survey company). Having providers be responsible for delivering the survey has obvious advantages for the LSC, in terms of resource, and reduced burden on the central administration function. This would also mean that providers had built in the desired satisfaction checks into their own internal systems. However, the associated risks are evident: providers may ‘cherry pick’ respondents, students may be pressured into responding positively; some learners may, inadvertently, be less represented (e.g. part-time learners, those learning outside the provider’s main location), therefore a stringent audit approach may be needed and more explicit guidance issued to providers. Ensuring sufficient rates of response for each provider to secure robust data is another challenge for the LSC. Although not an issue in the pilot learner survey, the pilot employer satisfaction survey did find different response rates and satisfaction levels depending on the survey method (telephone and web-based) and who undertook it (independent company and the provider itself).

Certain learners, particularly those with learners with learning difficulties and/or disabilities, are excluded from learner surveys at the provider’s discretion (although the LSC has commissioned work with learners at specialist colleges to explore this issue). Training providers may well collect the views of these learners but could include/exclude these views if they increase/decrease overall satisfaction ratings.

Survey content

In constructing the survey, it is evident that due consideration has been given to ensuring that the coverage of questions and the language employed is, as far as possible, accessible to all. It is not certain how much the extent to which the ordering of individual questions may influence responses to subsequent questions, although the

---

The fact that the ordering remains stable between providers does at least ensure that comparisons can be made between them.

The scoring of the survey, however, raises more intractable problems: the pilot learner satisfaction survey demonstrated that many learner responses clustered around the midpoint making it difficult to convert these scores to a four point rating scale on the PI. This means that it is difficult to develop banding scores (unsatisfactory, satisfactory, good, and outstanding) that effectively differentiate between these four categories. This issue is further compounded by sampling error and confidence intervals. Having confidence intervals of +/-3% within tight bandings can mean that many training providers can straddle the different bands. However, it is worth pointing out that one of the main reasons for the problem is likely to be the common tendency for respondents to avoid extreme responses to questions and to select the midpoints. Furthermore, there is considerable evidence that these tendencies towards midpoint response bias, or the opposite extreme response bias, are both culturally and socio-economically related (i.e., influenced by contextual issues). By deriving PI ratings simply by measuring the proportions of respondents falling into extreme categories (which may be influenced by contextual factors), the views of the majority of respondents are effectively seen as homogenous and consequently disregarded. It is clearly preferable to provide a more discriminating scoring scale than to resort to such measures.

The use of a seven-point scale may therefore be preferable to a five-point scale, although this is an issue that would require testing. While use of a three-point scale for learners with learning difficulties may facilitate the administration of the survey with such respondents, it further compounds the problem of lack of discrimination and makes comparison with other learners’ views difficult. Whatever the scale used – seven-point, five-point or three-point – it is prudent to use the same format for all questions on the survey to enable the score to be aggregated and converted to a rating with relative ease.

4.1.2 Performance Indicator: Learner Destinations

The Learner Destinations PI is essentially intended to determine the success of the learning provision in terms of improving the educational and/or employment circumstances of learners. It is a measure of the proportion of learners in one academic year who have progressed in the next academic year to one of five “positive” destinations such as a programme with the same or higher learning aim or a position of improved job security or career prospects.

The main issues here concern the accuracy and completeness of the data, the extent to which contextual factors, such as the surrounding labour market, are likely to influence the rating score, and the extent to which the five “positive” destinations can


be seen to be all-encompassing. For example, it is not clear whether a learner who goes on to a higher level of learning or more secure employment in a different country is included as a positive destination. Clearly, in terms of the individual, the destination is likely to be a positive one, but in terms of improvements to the local skills market, it is not.

From a statistical point of view, again the issue of homogeneity (lack of variation) in scores is a question of concern as this is likely to lead to particularly narrow percentage band-widths for the four different PI ratings. Where there is a high proportion of missing data (which appears to be particularly salient with this PI), the issue of confidence intervals becomes more prominent, with larger confidence intervals rendering the rating system at best liable to extreme year-on-year fluctuations and at worse meaningless. Perhaps, including within this measure, more variable outcomes such as average income of those going on to employment may widen the distribution of the data, but this considerably complicates the data collection process.

A more objective, accurate and complete method of collecting data might be by data matching (determining learner destinations from other objectively recorded databases such as HE destinations, tax records etc) and this is being trialled. This does, of course bring in concerns regarding access to such data, data sharing rights and that most thorny of issues in recent years: data protection.

4.2 Key Performance Area: Responsiveness to Employers

4.2.1 Performance Indicator: Employer Satisfaction Survey

Ensuring that a provider is responsive to the needs of employers (where relevant to the provider offer) is a key indicator of quality of the learning experience. Many of the issues concerning learner responsiveness are relevant here, including: the level of discrimination afforded by responses; using individual judgements as a basis for comparison across providers; and the operational issues around administering the survey and ensuring sufficient rates of response.

The pilot work also identified additional issues. The response rate was low, only about one quarter (26%) of eligible employers responded to the survey. Importantly, the response rate varied by the method in which the data was collected and by who conducted it. In short, response rates were highest for ‘phone surveys and those undertaken by a centrally administered survey. The LSC might trial different methods (including incentives) of increasing response rates.

In addition, levels of satisfaction also varied by who carried out the survey, with higher satisfaction levels recorded for those undertaken by training providers. In comparison with other surveys of employer satisfaction undertaken by independent research organisations employer satisfaction levels tended to be significantly lower in the FfE survey.8

As with the learner satisfaction survey, employer satisfaction ratings clustered around particular points raising the issue of how to place employers with similar satisfaction ratings in the different bands, especially when confidence intervals are included.

8 For example, Train to Gain Sweep 1 (http://readingroom.lsc.gov.uk/lsc/National/nat-tgemployerevaluation-may08.PDF) and National Employer Skill Survey 2007 (http://readingroom.lsc.gov.uk/lsc/National/nat-nessurvey2007mainreport-may08.pdf)
For example, the Employer Satisfaction Survey must contend with similar concerns over the ability of the scoring methodology to discriminate over a range of satisfaction levels, although it is at least relieved of the issue of needing to adapt the survey to those with learning disabilities.

Furthermore, as with the Learner Views Survey, there is a problematic issue of representativeness which is further compounded by far lower response rates. Whereas for the former, the samples of learners represented a significant proportion of learners at each provider, for the Employer Satisfaction Survey, the response rates in the pilot survey averaged only 22 per cent, which despite the contracting of a professional survey company to manage future data collection, is likely to be an ongoing problem. This, and the fact that the response rates will vary depending on the method by which employers will be approached, will have a much more of an impact on representativeness, not only because confidence intervals are likely to be considerably larger (rendering the measure less robust), but also because there is likely to be a considerable response bias. In other words, it may well be that those employers who do respond are more likely to be those that hold stronger views (or extreme levels of satisfaction). The extent to which the obtained data can be generalised to each provider's entire population of employers is consequently likely to be highly limited.

All of the general concerns with regard to surveys (see section 3.3 under cross-cutting issues) will have greater 'bite' with this survey, due to the low numbers and percentage of respondents.

4.2.2 Performance Indicator: Amount of training

This measure is based on the volume of learners on "key programmes", weighted by provider size. In theory then, it is based on objective data obtained from ILR records and, so long as that data is complete, accurate and up-to-date, and identification of "key programmes" is appropriate (i.e. linked to an evidence base detailing what learners and employers want in terms of responsive provision), there should be few specific concerns over statistical robustness.

In developing this PI, the LSC has consulted with providers, in order to make sure it adopts the right approach as far as they are concerned. A key priority was to ensure rewarding growth and avoiding penalising stability in terms of levels of provision. The ongoing development (which moved from the original focus of fees and volumes of training delivered towards an amount of training measure) is to encourage providers to develop into new areas, expanding their training offer and bringing in new learners, and making new links with employers.

Where problems arise for this (and the Employer Satisfaction Survey) PI they may primarily be concerned with the inclusion of the Training Quality Standard (TQS) Accreditation: those providers which are judged to meet criteria concerning the way in which they meet employers' needs, will be accredited with the TQS and will automatically receive the highest rating across both PIs for the Responsiveness to Employers KPA for three years.

However, these judgements will be made on the basis of the TQS Assessment Framework, assessing how far the provider is responding to the needs of employers as individual customers, and how far the provider is developing and deploying products to
address particular sector needs. Exactly how far the criteria of these judgements align with the FfE PIs, how the judgements will be independently verified and quality assured and whether the TQS can be seen as a proxy for the FfE Employer Satisfaction Survey and the Amount of Training PIs are key questions. In other words, the TQS represents a dual measurement system whereby providers may bypass the usual measurements of the PIs in this KPA and instead obtain a TQS accreditation which is assumed to be equivalent. Furthermore, the TQS assessment is made 3-yearly while the Employer Satisfaction Survey and the Amount of Training analyses are conducted yearly.

It is worth stating, however, that our impression of the TQS, is that it is based on a relatively robust rationale and practice. Nevertheless, using accreditation with the TQS as a proxy measure of quality across the whole of the employer responsiveness represents a relatively unusual approach. The LSC needs to be comfortable that the TQS represents an adequate substitute for the employer responsiveness PIs i.e. employer satisfaction survey and amount of training. As far as possible, it should be explicit that the TQS as a proxy is replacing ‘like for like’ in terms of quality and performance. It would also be necessary to determine the level of risk in terms of the potential to drive perverse behaviour in providers if they are moving towards TQS accreditation, if they see this as an alternative to succeeding against the stated employer responsiveness PIs.

Until these issues are addressed, it might be prudent to collect data (though not publish it), for all providers accredited with the TQS, concerning what their rating on each PI would be had they been assessed using the usual measurement tools and also what PI ratings those with poor TQS ratings achieve. Analysis of such data would provide a clearer picture of the parity of the measures and might also be able to determine any year-on-year fluctuations in PI scores that are not reflected in the fixed 3-year TQS rating.
5 EFFECTIVENESS DIMENSION

5.1 Key Performance Area: Quality of Outcomes

5.1.1 Performance Indicator: Qualification Success Rates

The Qualification Success Rates PI is the sole measure of Quality of Outcomes, although as discussed previously, an argument could be made for the inclusion of the Learner Destinations PI within this Key Performance Area.

Although the PI is essentially an empirically derived measure, several questions arise:

- Is the derivation of the rating from the raw data based on a clear rationale?
- To what extent does the PI actually reflect success?
- Does the Quality of Outcomes rating adequately compensate for input factors (contextual factors) at individual providers (i.e., to what extent is it a measure of value-added as opposed to pure outcome)?

Derivation of the PI rating

This is worth exploring in some detail given that, at first glance, the method by which the rating is obtained is not obvious. Nevertheless, the process by which the Qualification Success Rates rating for providers is derived does follow a fairly clear rationale:

a) On the basis that different qualifications are achieved with different relative ease or difficulty, the qualifications themselves are firstly divided, broadly, into five categories: FE long courses, FE short courses, GCEs (AS- and A-levels), Apprenticeships (including Advanced Apprenticeships) and Train to Gain qualifications.

b) Within each provider, of those learners that started the course, the proportion who achieved a pass in each qualification category is obtained (from ILR and LAT data) and is compared to the national average for the qualification category. Depending on which percentile of the national distribution this falls into, this percentage is allocated a point score. For all qualifications except for GCEs, the maximum number of points available is 130.

c) As a response to wide-spread consultations, for GCEs, the maximum number of points available from this proportional success rate is only 100. However, up to an additional 30 points are available for value added. Here, for each provider, the difference in prior attainment of learners (based mostly on GCSEs and other level 2 qualifications) and final GCE attainment (which is essentially based on grades) is calculated and compared to the national average. If the value added is significantly lower than the national average, no additional points are awarded. If it is not significantly different from the national average, 15 points are awarded and if it is significantly greater than the national average, then 30 additional points are awarded.
d) The point score for each qualification category within a provider is then weighted according to the proportion of learners at the provider who took the qualification. So, if the point score for FE long courses at a provider were the full 130 points (i.e. all learners who started the courses achieved passes), but only 1 per cent of learners took the course, then the provider would (rightly) only obtain 1.3 points for FE long courses.

e) The weighted point scores for each qualification category is then summed and, depending on which quartile the aggregated point score falls into, the provider is allocated a rating of 1 (Outstanding) to 4 (Inadequate).

One potential problem with the process may lie in step c), where the proportion of learners who pass the GCEs are compared to the national average for attainment at Sixth Form Colleges (SFCs) and not to the average attainment at all providers. The rationale for this is that, in practice, most GCE courses are undertaken by learners at Sixth Form Colleges, so to facilitate like-with-like comparisons it is the national average among such providers that is used as the benchmark. While this does enable the majority of GCE providers (SFCs) to obtain a fair measure of Qualification Success Rate, it rather disadvantages other providers (such as FE colleges) in which the national average pass percentages (which would be the appropriate benchmark for such providers) are actually substantially lower.

The FfE consciously avoids separating out PI ratings by type of provider, not least because provider-type categorisations tend to fluctuate a great deal. One could argue that if it could be shown that the SFC/non-SFC distinction was a relatively stable one, a better method might be the following: continue comparing the proportion of learners with GCE passes at SFCs with the SFC national average, but compare the attainment at all other providers with the national average across all providers. This would at least diminish the disadvantage suffered by non SFCs, albeit a relatively small problem given the small numbers of GCE courses taken outside SFCs. Evidence from more complete data, however, suggests that the GCE grid as it stands, has in fact been optimised for fairness to different types of providers and, as stated, consultation with providers shows the decision to base the national average benchmark for GCE on that in SFCs to be well supported.

An accurate reflection of success?

The empirical approach employed for this PI clearly gives it the advantage of being objective and it does give some indication of success, albeit a crude measure based largely on pass-rates. Where it fails short is in discriminating between levels of success. For example, while many FE short courses are graded dichotomously (i.e. pass or fail), some FE short courses and a greater number of FE long courses are graded more discriminatively using categories of fail, pass, merit, or distinction (or similar). For current GCEs the grading system is even more discriminating with passing grades ranging from A through to E, “fails” being ungraded (U) and the new GCEs (available for first teaching from September 2008) due to be yet further discriminating by inclusion of the A* grade at A-level. The simple measure of

---


proportion of learners passing these qualifications does not take account of the considerable variability in success that a pass encompasses. One cannot necessarily assume that providers with success rates based purely on pass proportions would, relative to each other, score equivalently if success rates were based on more discrete measures of success such as grades awarded. While the grading of GCEs is tangentially considered in the calculation of the value added component, as discussed in the following section on input factors, it accounts for only a small proportion of the calculation of this PI.

With schools tending to take an ever increasing interest in grades as opposed to pure pass-rates, we would recommend that some means of incorporating the proportion of learners achieving particular grades at GCE, rather than the proportion achieving passes, should be incorporated into this PI.

**Input factors:**

A key question with regard to the utility of this PI, particularly whether it is able to allow fair comparisons between providers, is whether it takes into account baseline input measures such as the prior level of attainment of learners on courses. Clearly, those providers with highly able, high achieving learners, will more easily be able to attain high rates of qualification success, without necessarily having to deliver high quality provision and support, and vice versa.

Arguably therefore, it is in relation to this PI that contextual factors such as relative deprivation, ethnic learner groups and rurality, which are likely to influence the prior attainment levels of learners, might have the biggest impact. This is a potentially a serious shortcoming in the use of this PI, and may represent a far greater disparity between actual and perceived Quality of Output than the issue of hypothetical super-populations which appears to have dominated discussion of the PIs in general. One of the most important practical implications of the shortcoming may be to increase the risk of “perverse behaviour”. The main concern is that it would be in the provider’s interest not to recruit certain learners who are less likely to achieve. This would obviously run contrary to the LSC’s equality and diversity, and widening participation priorities.

The LSC is looking at resolving this issue. Furthermore, it should be noted that while at a conceptual level one can understand the concern, early indications are that the first step in calculating the PI – division by qualification type – greatly diminishes contextual influences, no doubt because choice of qualification type is itself influenced by contextual factors.

5.2 Key Performance Area: Quality of Provision

5.2.1 Performance Indicator: Inspection Grade

The Ofsted inspection grade awarded to each provider will be included in FfE data, without exception. FfE scores will be derived from the most recent Ofsted grade awarded based on the judgements of inspectors against the Common Inspection Framework.

Making use of Ofsted inspection information ensures that FfE effectively links with other established sources of data, utilising existing judgements and quality assessment results to offer a richer, more rounded picture of provision.
**Issues arising**

A small number of issues have been identified associated with this PI. The main one of these is the frequency of Ofsted inspections and more broadly the level of synchronicity between the production of FfE scores and Ofsted grades resulting from inspection. Current Ofsted inspection cycles mean that a provider can be waiting for up to three years for an inspection, and discussions are ongoing to increase the period between inspections for high performing providers to six years. Therefore, given the annual cycle of FfE scores, there is potential for lag which may affect the accuracy of information.

For colleges or providers which are new or have not been subject to Ofsted inspection, these are exempt from this key performance area until an inspection has been carried out.

In addition, stakeholders raised the issue of potential differences between the Ofsted inspection judgement and the FfE rating of providers. As both frameworks are measuring and assessing quality, this should not be a common problem. However, should this arise, this may be due to differences in the reporting cycles (as mentioned above), or perhaps some of the different areas covered by the systems (e.g. Ofsted inspections including judgements on the leadership and management of a provider).

The LSC will need to be prepared to directly address concerns raised by providers in relation to both FfE and Ofsted ratings, and to explore contributory factors in cases where these are mismatched.
6 FINANCIAL DIMENSION

6.1 Key Performance Area: Financial Health

As with all of the following financial indicators, the Financial Health KPA is a purely empirical measure based on what appear to be sound rationales. Specifically, the Financial Health KPA comprises three PIs:

- current ratio (solvency)
- operating surplus or deficit as a percentage of turnover/income (sustainability)
- borrowing as a percentage of certain reserves and debt (status).

A concern expressed with the Financial Health KPA was that it was predicated on the FE College model where such a measure is much more straightforward and clear to develop compared to a private or work based learning provider. The issue around burden of reporting has been raised previously, see section 2.2, and the applicability of PIs across provider types. In practical terms this issue can have significant impacts on providers who are required to provide information or data they do not already have.

There were also concerns raised about ‘compromises’ on capital build which mostly affects FE Colleges and so distorts a comparison between training providers. It was also felt that some FE providers were concerned about the ways in which the capital build issue had been addressed. However, there is a facility for the LSC to moderate the grade depending on individual provider circumstances.

6.2 Key Performance Area: Financial Control

The Financial Control KPA also appears to be predicated on a college model. However, unlike the Financial Health KPA, it is based on self assessments by the providers, some of which are validated by the professional judgement of auditors such that over the course of three years, all providers are likely to have been audited.

6.3 Key Performance Area: Use of Resources

The Use of Resources KPA, as with the Financial Health KPA involves a composite measure although of considerable more complexity: It comprises three Performance Indicators as follows:

- Funding Economy, itself comprising two measures:
  - Use of Resources 1: “the proportion of LSC funding applied to priority provision”.
  - Use of Resources 2: “delivery as a percentage of funding allocation or contract value”.

- Resource Efficiency, again comprising two measure:
  - Use of Resources 3: “LSC funding per successful outcome”.
  - Use of Resources 4: “provider level unit cost”.

<table>
<thead>
<tr>
<th>G</th>
<th>H</th>
<th>K</th>
</tr>
</thead>
</table>
• Capital, comprising one measure:
  o Use of Resources 5, which itself consists of two measures:
    ▪ condition: the current condition of the college’s building stock (based on the judgements of Regional Property Advisors (RPAs));
    ▪ renewal factor: which takes into account the current condition.

Several issues arise in relation to this KPA:

• Due to exemption criteria, full data is available only for a small number of colleges. However, three out of five of the Use of Resource measures (above) can be calculated for all providers, enabling some comparison between providers.

• The differences between different points scores was slight e.g. 0.5% (Use of Resources 2) or as little as £40 (Use of Resources 3). Whilst this may not be significant overall, there is a general issue of the use of composite indicators (see below and section 3.3).

• The ability of training providers to use the FfE in their own performance management structures. On these PIs, there is a lot of potential to help training providers compare their costs of delivery of successful outcomes, but most of this data is only available for colleges.

6.4 General issues concerning the financial dimension

An important element of the finance dimension was that it was predicated to the FE model and that data would not be available for other types of provider or it would be more problematic to collect. There are additional considerations over the use of professional judgements (including those of Regional Property Advisors) and a 3-4 year cycle of monitoring. In addition, there are different criteria for different types of providers and, with slight differences in the points scoring mechanism could accumulate to unnecessarily advantage or disadvantage providers.

One final point, is that of the use of composite measures (further explored in section 3.4). With each of these three financial KPAs comprising several PIs, which in turn usually comprise several different measures, there is a great deal of raw data being aggregated into a small number of ratings. The difficulty here, is the extent to which individual components are first converted to ratings (or similarly simplified) before being aggregated, leading to small errors being compounded on an aggregated scale and providers who narrowly miss out on particular ratings, being significantly advantaged or disadvantaged at a holistic level. Similarly, slight changes in the decision rules and weightings used can have a considerably larger global impact on training providers’ overall scores. The fact that the financial dimension in particular is a composite measure consisting of so many sub-measurements renders it more prone to these problems than other KPAs.
CONCLUSIONS AND RECOMMENDATIONS

7.1 Overview

Broadly, the FfE can be deemed to be successful against its stated aims and objectives, as set out in key documents:

“The main aims of the Framework are to provide a single, unified framework for assessing and reporting achievement in all key areas of performance. The use of the Framework will lead to better and more relevant outcomes for learners and employers, and a better reputation for the FE system” (Framework for Excellence: Raising Standards and Informing Choice, LSC, March 2007).

In terms of improving outcomes and enhancing reputations, the impact of FfE will not become apparent until further down the line. However, the achievement of a single unified framework against which performance can be reported and assessed is evident.

The principles of design, development and implementation (set out in Annex B), have successfully guided the evolution of FfE and adhering to these will ensure that the ongoing development of FfE will continue to address the identified priorities for activity.

In terms of the three main areas for consideration outlined in section 1, and based on the research activities undertaken as part of this Review, we can conclude that:

- FfE is measuring the right areas in order to present an accurate picture of provider quality and performance: for example, FfE gives a relatively high weighting to responsiveness, and ensures that learner and employer experiences inform quality judgements. It is possible that, as the Framework evolves, the LSC may consider incorporating other dimensions of PIs, such as Responsiveness to Communities, with reference to the Audit Commission approach to Comprehensive Area Assessments, which adopts an holistic approach and looks at services within an identified area in the round and not at specific service providers;

- FfE is measuring things in the right way: our Comparative Review found that FfE compares favourably in relation to other existing systems, for example in the levels of transparency and coverage of performance areas. The FfE should provide a tool to support provider self-improvement, although initially larger, more established providers will most likely be in a better position to understand and use the data. However, there are areas for further consideration, such as the combination rules, and issues around contextualisation which still need to be resolved in order for FfE to present a final set of information for a specific provider;

- FfE data is generally being interpreted in the most appropriate way, as described in section 3: the Review of individual PIs highlighted the strengths and risks associated with the individual measures and underpinning elements. The LSC has taken thorough steps through consultation, discussion and
targeted research, and informed development (e.g. through the Technical Working Group, and Policy Committee) to make sure that considered decisions are made and the associated potential impacts (both negative and positive) are highlighted. Attention to outstanding issues, such as those highlighted below, should continue in this vein.

As stated above, at the time of writing, some areas relating to FfE are yet to be finalised, such as the rules governing combination of scoring, and others which represent new areas of activity, such as the employer survey. These may represent challenges to FfE in terms of achieving a final, rounded approach. However, the LSC is aware of the issues relating to individual PIs and the development of underpinning approaches and methodologies is being undertaken according to stated approaches and timescales.

7.2 Issues to be considered and/or addressed

The study identified a set of issues of which the LSC need to be aware, and some which also necessitate action to resolve them. These relate principally to the stakeholder consultation, cross-cutting issues identified in section 3 and the individual PIs reviewed in sections 4, 5 and 6, and are set out below:

7.2.1 Issues identified by stakeholders

- Validity and reliability of data – this relates principally to the use of surveys to generate scores e.g. for responsiveness to employers and learner satisfaction. The approach to sampling and the administration of the survey have both been identified as areas for concern through the survey. As discussed in section 3, the use of surveys presents inherent issues of reliability and robustness. On balance, direct consultation with key users of provision is a highly desirable aspect of performance measurement. The risks associated with survey use are well documented, and the LSC is aware of them. Questions also arose regarding the value of FfE data to different user groups, and ensuring that the outward facing aspects of FfE serve their intended purpose. Research is currently being undertaken to identify user groups, to scope their data and information needs and to ensure that FfE meets these needs.

- Coverage of FfE – the FfE, both in terms of the Key Performance Areas and the individual PIs, covers the key areas that one would expect to see in a framework such as this. Stakeholder feedback suggested that additional areas e.g. responsiveness to communities, or partnership working would also be helpful in determining the quality of a provider. The early phase of development of FfE dictates that simple is best at this stage, and the LSC are to consider additional areas of measuring performance in the future. FfE coverage compares well with other performance measurement frameworks. Incorporating ‘user views’, whilst not without its problems, is a distinct asset for FfE.

- Focussing on measures, and responding to reporting requirements – concerns were voiced by a small proportion of stakeholders that perverse behaviour could result in some providers, as a result of FfE. This was not a significant concern, and the general consensus was that, the vast majority of providers are delivering to a certain standard and therefore would not need to resort to ‘gaming’ or attempting to manipulate scoring in some way. The use of
absolute measures and clearly stated PIs, combined in a transparent way also reduce the risk of perverse behaviour being driven by FfE. Moreover, as FfE is making use of a significant amount of existing data, the opportunity and also the motivation to manipulate resulting data will be limited.

7.2.2 Issues relating to cross-cutting themes

This section provides key considerations in tackling problems with associated cross-cutting themes. Many of these considerations have previously been addressed or are in the process of being addressed within the Framework. Where they have not been, however, they constitute important recommendations. Of course, some of the considerations will not require any active practices to be implemented, but merely an awareness of their existence, and possibly guidance to those interpreting the Framework.

Balance of weighting:

- Issues:

  1. The FfE has a broad user base which is reflected in a broad measurement base in which dimensions are given equal weighting, which may or may not be an accurate or correct reflection of the relative importance of different dimensions on provider performance.

  2. Providers for whom certain KPA data are not collected may be unfairly advantaged or disadvantaged because of the increased relative importance of the other dimensions.

- Recommendations:

  1. FfE should provide guidance in the interpretation of OPRs where providers are exempted from an entire KPA.

Contextualisation: impact of contextual factors (e.g. ethnicity, local educational attainment, population density etc.) on performance scores.

- Issues:

  1. Initial findings of analyses suggest there are no clear significant influences of contextual factors on FfE KPAs or PIs, this is based on limited data available from the pilot.

- Recommendations: depending on evidence from analyses of complete data, it may be necessary to implement one or more of the following actions:

  1. Exclusion of PIs heavily influenced by context;

  2. Adaptation of PIs to include contextual value added (CVA);

  3. Provision of guidance to the interpretation of PIs most affected.
Use of surveys:

- Issues:
  1. Minimum number of responses should be set for providers;
  2. Survey of smaller providers should be based on complete census rather than selective sampling;
  3. Method should be prescribed in detail, including method of sampling and dates for completion;
  4. Responses should be weighted from each provider to ensure they are representative of the population;
  5. Benchmarks should be set to distinguish levels of performance;
  6. Additional questions should be available from a ‘bank’ so that providers can ‘piggy back’ them on their own surveys.

- Recommendations: none as issues have been addressed, although in the case of issue 6 (‘piggy backing’ onto providers’ own surveys), the action has been deemed unworkable, at least in the near future.

Use of composite measures:

- Issues:
  1. Aggregation of individual performance measures may disguise serious failings in some or part of the system;
  2. Combining PIs of variable quality essentially reduces the composite measure to the lowest common denominator;
  3. Excluding poorer PIs can distort performance scores;
  4. Composite measures are highly sensitive to weightings, which should therefore be formulated with clear rationale;
  5. Relatively few PIs should be based on sampling.

- Recommendations: no specific recommendations as the FfE already has comparatively few PIs and KPAs and has developed them with cognisance of the issues.

Combination rules:

- Issues:
  1. Combining point scores is preferable to combining weighted grades, as it allows greater discrimination in OPR grades.
2. Regression to the mean, and combining of some PIs with very narrow distributions results in OPR ratings clustering towards central values (of 2 and 3).

- **Recommendations:** none as the FfE has adopted measures to deal with both issues, and intends to review these in the light of complete data sets.

**De minimis criteria and exemptions:**

- **Issues:**
  1. FfE is viewed by some as being largely predicated on an FE College model; other training providers are quite different, and the level and quality of their data is not necessarily as robust.

  2. Weighting is crucial, as is ensuring that exemption levels do not affect the overall balance of results in terms of offering advantage or disadvantage to providers.

  3. *De minimis* rules may encourage some training providers to reduce capacity below exemption levels and so avoid having to produce certain PIs. This relates back to the wider issue of driving perverse provider behaviours, discussed in more detail elsewhere in the report.

- **Recommendations:** none, as the new combination rules reduce the impact of all of these issues to negligible levels.

**7.2.3 Issues relating to individual PIs**

This section summarises key recommendations, or issues for consideration and awareness for the individual KPAs and PIs. Where a KPA/PI is not included below it is because there are no specific issues arising from it.

**Key Performance Area: Responsiveness to Learners**

- **Performance Indicator: Learner Views Survey**
  - Awareness is needed of the existence of midpoint response biases and extreme response biases and their tendency to be related to socio-economic factors.
  - Attempts to provide a more discriminating scoring method, perhaps by using a wider point scale, should be investigated.
  - Whatever the scale used, it is recommended that a consistent scale is used throughout the survey, as much as is feasible, in order to facilitate aggregation of scores.

- **Performance Indicator: Learner Destinations**
  - The possibility for collecting data by data matching should be investigated.
**Key Performance Area: Responsiveness to Employers**

- **Performance Indicator: Employer Satisfaction Survey**
  
  In addition to consideration of the same issues that impacted on the Learner Views Survey, awareness is needed that all of the general concerns with regard to surveys (see cross-cutting themes) will have greater ‘bite’ with this survey, due to the low numbers and percentage of respondents.

- **Performance Indicator: Amount of training**
  
  - it might be prudent to collect data (though not publish it), for all providers accredited with the TQS, concerning what their rating on each PI would be had they been assessed using the usual measurement tools and also what PI ratings those with poor TQS ratings achieve.
  
  - It is also necessary to determine the level of risk in terms of the potential to drive perverse behaviour in providers if they see TQS accreditation as an alternative to succeeding against the stated employer responsiveness PIs.

**Key Performance Area: Quality of Outcomes**

- **Performance Indicator: Qualification Success Rates**
  
  - We would recommend that efforts to incorporate the proportion of learners achieving particular grades at GCE (and in other qualifications), rather than the proportion achieving passes, should be investigated.
  
  - Adjustments for contextual factors or the extent to which calculating the PI by qualification type ameliorates this problem should be investigated.

**Key Performance Area: Quality of Provision**

- **Performance Indicator: Inspection Grade**
  
  The LSC will need to be prepared to directly address concerns raised by providers in relation to both FfE and Ofsted ratings, and to explore contributory factors in cases where these are mismatched.

**Key Performance Area: Use of Resources**

Several issues arise in relation to this KPA:

- Due to exemption criteria, full data is available only for a small number of colleges. Consequently, it may not be appropriate to compare one provider with another on the basis of this KPA.

- The ability of training providers to use the FfE in their own performance management structures. On these PIs, there is a lot of potential to help
training providers compare their costs of delivery of successful outcomes, but most of this data is only available for colleges.

**General issues concerning the finance dimension**

- There are different criteria for different types of providers and, with slight differences in the points scoring mechanism, these could accumulate to unnecessarily advantage or disadvantage providers. The extent to which this is a problem requires investigation although it may be that no action should be taken as a result: merely an awareness that there is an issue here.

- There is a great deal of raw data being aggregated into one a small number of ratings. As far as possible, individual components should be aggregated in ‘raw’ form before conversion to a simplified rating, rather than vice versa, to avoid small errors being compounded.

**7.3 The overall ‘health’ of FfE**

To all intents and purposes, it would appear that throughout the process of development underpinning FfE, the LSC and partners have given due consideration to the relevant issues and factors which may have an impact. The evidence suggests that the level of consultation, research, discussion and debate, as well as technical knowledge employed, has resulted in a Framework which offers a relatively robust picture of provider quality and performance. As highlighted, there are specific issues which relate to individual PIs. However the LSC is aware of the potential risks of these, such as survey representativeness, use of proxy measures such as the TQS, and debates around the qualification success rates measure. The LSC has undertaken research and consultation exercises to try and address these, and the evidence would suggest, has thus far been able to balance the demands of FfE and the related need to gather and present useful, defensible data, with the need to be pragmatic in terms of decisions concerning the parameters, limitations and caveats that apply to some data.

On a final note, following the launch of Version 1 of Framework for Excellence in June 2008, the next phase of activity, and a key challenge for the LSC and partners, is to ensure that the data and information gathered and presented under the FfE banner, is made available in a format that is accessible and appropriate to the needs of identified user groups. The LSC are currently addressing the ‘outward-facing issues’ of FfE data through the development of the Communications and Marketing Strategy for FfE. An integral part of this will be ensuring that FfE is used in the most fruitful way in order to drive up quality and standards, inform learner and employer choices, and serve as a self-improvement tool across the provider population.
ANNEX A – KEY RESEARCH QUESTIONS

The original Invitation to Tender produced by the LSC for the Independent Review detailed a number of key research questions guiding the study:

a. Does FfE measure provider performance in the right areas, with the right indicators, or are there any obvious gaps in the framework? Are the FfE indicators and data sources and associated assessment criteria defined appropriately?

b. Given inherent differences between different types of provider, is it appropriate to measure performance using a generic set of indicators?

c. Are the FfE data collection methods and sources robust and appropriate, in particular for survey data?

d. How appropriate/robust is the use of a composite measure to summarise overall performance? Does FfE have a sound and appropriate approach to combine individual indicator scores into scores/grades for the whole provider?

e. Should differential weighting be applied in the construction of such composite measures; and to what extent are they sensitive to different weighting or other methodological choices/changes over time?

f. Are differences in both the individual and the composite indicators likely to reflect real variations in performance?

g. Is FfE consistent with established performance management practice/principles for FE colleges/providers, and more generally; and, in the light of this, is FfE likely to be an effective management tool for providers?

h. To what extent are FfE indicators likely to drive unforeseen perverse behaviours in the provider population?

The Review may also consider the following questions; but these are lower-priority, partly because information in these areas will be available from other sources.

i. To what extent will FfE support greater and more informed learner and employer choice?

j. How can we make the Framework sufficiently accessible as to be easily digested by providers or lay users, or is it overly complex?

k. Identification of risks, and proposals for managing these
ANNEX B – PRINCIPLES OF DESIGN, DEVELOPMENT AND IMPLEMENTATION

Design

1

A number of principles underpin the design of the Framework. These are that the Framework should:

• have as its core purpose serving the needs of customers, that is, learners and employers. That means giving priority to the views of learners and employers in the event of a major divergence of opinion between stakeholders

• be clear, simple and transparent so that performance information is easily available for users, in forms they find helpful

• be valid, providing information that is a robust and reliable guide to the major aspects of an institution’s performance

• build on the Common Inspection Framework and New Measures of Success

• be based on absolute rather than relative performance measures

• be fair, in that it does not unduly advantage or disadvantage any organisation or type of organisation

• be transparent in the way that indicators and ratings are chosen and constructed.

2

Performance indicators and the definitions of key performance areas should:

• be fit for their diverse purposes, including informing learner and employer choice, organisational management, quality improvement and external accountability

• be capable of ‘drill down’ to produce detailed performance information to inform learner and employer choice

• where new, not be included without testing and piloting and the advice of the Information Authority and the Bureaucracy Reduction Group.

Development

3

The principles underpinning the development process of the Framework are that it should be:
• open and transparent, with the publication of regular progress reports

• based on continuing dialogue and consultation with providers and learners and employers.

**Implementation**

4

Implementation of the Framework will be undertaken according to the following principle.

• Framework scores and ratings should be available in a public form to learners, employers and communities.