Review of Data Systems Underpinning DfES SR2004 PSA Targets

by

Ian Schagen

March 2006

Review of Data Systems Underpinning DfES 2004 PSA Targets

Contents

Appendices

1 Introduction

Public Service Agreement (PSA) Targets have been in existence since 1998 and are agreed by the Department as part of the Spending Review process. The National Audit Office (NAO) reviewed and validated the data systems underpinning the SR2002 targets in 2004 and the SR2004 target systems in late 2005 and early 2006. As part of the Department's input to both validations, and to gather information to help in the development of its data systems, an external consultant was commissioned to coordinate a team of the Department's statisticians in the task of gathering information and making judgements about each relevant data system. This document should therefore be regarded as an independent report to the Department, based largely on work carried out by departmental and NAO staff and others. The individual data system reviews on which this document is based also underpin the report which is being produced by the NAO on the SR2004 PSA target data systems[1].

The aim of the review was to:

1. Assess each of the Department's live PSA targets against the NAO's validation approach.
2. Set out recommendations for how the underlying data systems should be enhanced / improved.
3. Set out recommendations on how data issues should be addressed in DfES for the next round of PSA targets in the 2007 Comprehensive Spending Review.

Each statistician worked closely with a member of NAO staff in carrying out the review. Where a target in SR2004 was carried forward from SR2002, then the two-person team worked on both versions, addressing issues raised in the initial review of the SR2002 targets for the latter and conducting a thorough review from first principles for the former. Details of the procedures used in the SR2004 reviews are given in the next section.

This comprehensive report is intended to fulfil a dual function: (1) It summarises the work that was undertaken in support of the NAO's overall review and validation of the SR2004 targets, (2) It is intended to be a source document from which the Department can continue to examine and improve its data systems, not only for measuring progress towards PSA targets, but for a whole range of other purposes.

Appendix 4 contains a summary of the review of the SR2002 target data systems; some of the issues flagged up in that review carry forward into the present work.

[1] The joint PSA 2004 target on Young People and Drugs will be reviewed as part of the validation of Home Office targets.

2 Description of the review process

The review process was conducted almost entirely by a set of 14 two-person teams, each comprising a Department statistician and an NAO examiner. Central direction and coordination was carried out by an external consultant for technical issues, with administrative and technical support provided by another Department statistician. A project steering group, chaired by the DfES Head of Profession for Statistics, met at regular intervals to provide strategic advice.

Because the data systems were being reviewed from both a statistical and a risks and controls standpoint, it was important to pull together these two different viewpoints and produce a process for review and reporting which was coherent but enabled all appropriate judgments to be captured. At the end of the review process, each review team agreed a validation rating for the data system, based on NAO's criteria for such ratings, set out in Figure 1 below.

Two generic documents were produced, reviewed and finalised before being sent to each of the 14 teams to help in their deliberations:

1. A **statistical checklist** (see Appendix 1). This was intended to take the team through a series of questions from a statistical viewpoint, trying to uncover risks to the validity of the data system in a systematic fashion.
2. A **reporting template**. This working document was based heavily on NAO's standard template for reporting the outcomes of its validations of systems, slightly modified to include additional features such as a specific section on the Technical Note.

To enable both of these guidance elements to be used effectively, it was important to set out the relationships between them, so links were embedded in each document to demonstrate how judgements made in one related to elements of the other.

These two documents indicate slight differences in emphasis between the statistical and the NAO approaches to the review. To a large extent these are matters of terminology, and both approaches were designed to answer the same basic question: "Is this data system fit for the purpose of measuring whether the underlying target has been met?". The statistical approach starts by defining the underlying population of interest, and then whether data has been collected from it in a representative manner and whether the measurements taken are valid and reliable. It is concerned about bias and uncertainty in estimates. The NAO approach takes the same considerations into account, using the terminology of 'risk' and 'control', and also focuses on verifiable processes and adequate disclosure of potential problems. The 'two-pronged' review system provided a powerful and collaborative way of determining the strengths and weaknesses of the data systems.

The criteria for the reporting template were arranged under five main headings:

1. Specification
2. Operation
3. Reporting
4. Technical Note

5. Management controls

Within these headings, the specific criteria were:

1.1 Is the system ***relevant***? (Does it cover all significant aspects of the target)
1.2 Is the system ***well-defined***? (Unambiguous definition, collected consistently)
2.1 Is the system ***robust?*** (Sound procedures for identifying risks and effective controls to address them)
2.2 Is the system ***reliable?*** (Accurate enough for the required purpose)
2.3 Is the system ***comparable***? (With past periods and for detecting change over time)
2.4 Is the system ***verifiable?*** (With clear documentation and processes which can be validated)
3.1 Does the system ensure ***clear*** reporting? (Unambiguous about what is and is not being measured)
3.2 Does the system ensure ***transparent*** reporting? (No concealed manipulation of data)
3.3 Does the system ensure ***comprehensive*** reporting? (All aspects of data system relevant to target are included)
4. ***Technical Note***
5 ***Management Controls*** (Consider the adequacy of the review and validation procedures leading up to the adoption and use of the data systems and the regular reporting on them).

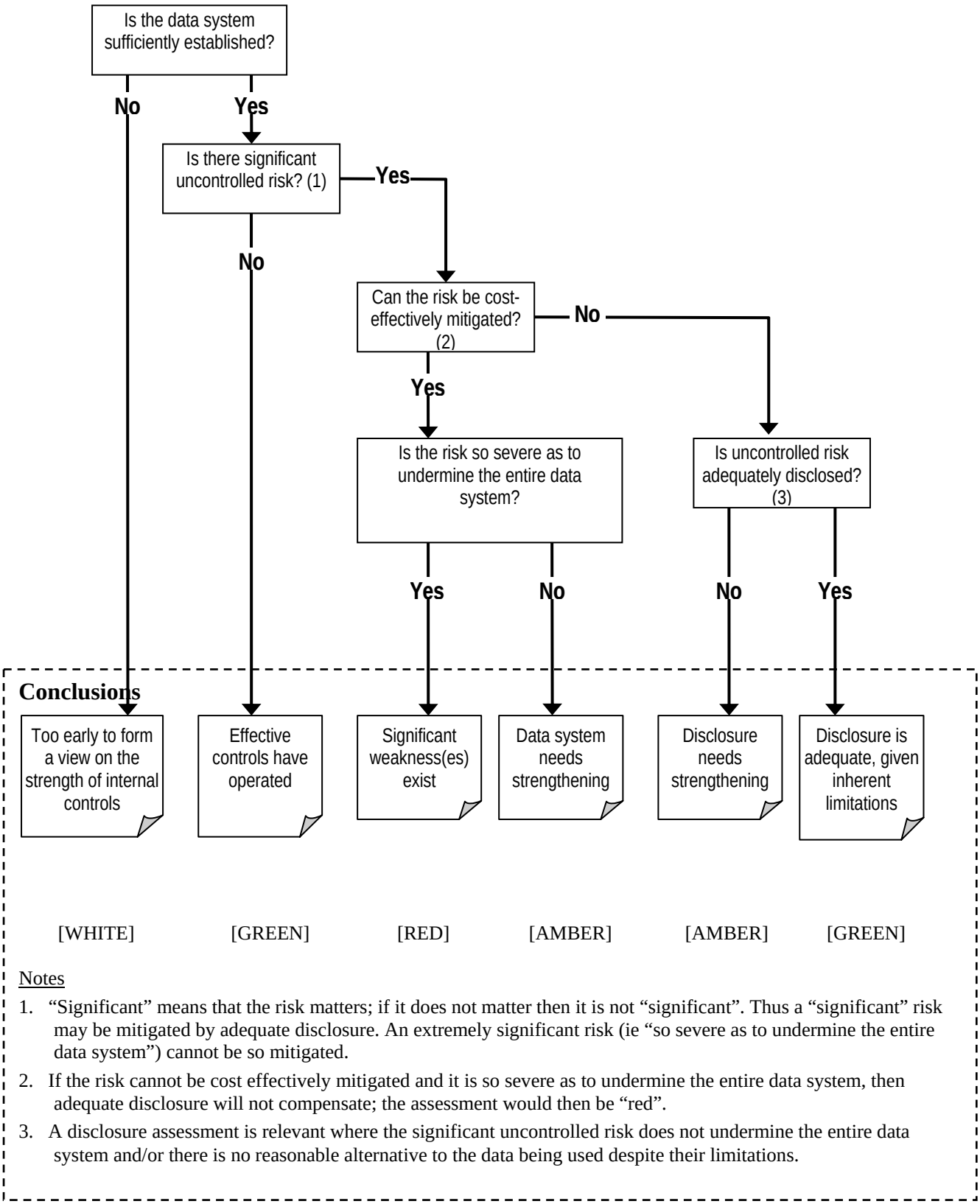
In the process of the review, each team completed a reporting template in the light of:

- Published documentation
- Discussions with the data owner (in the Department or other government department)
- Discussions with the Department policy lead
- Discussions with other relevant individuals, including those responsible for day to day operation of the data system, and staff in other organisations with a significant input.

Initial versions of the reporting template were sent to the external consultant, who carried out a certain amount of editing to ensure consistency of style and returned them with comments, queries and recommended further work. This process was repeated once or twice for each data system, until there was agreement that all aspects of the data system had been rigorously evaluated. Personal interviews between the external consultant and each statistician were conducted on two occasions, in order to raise significant issues about each data system.

The review and consequent recommendations for each target were summarised by the external consultant, and were checked by data owners and policy teams for factual accuracy. Summaries of reviews and recommendations for each target are included in this report at section 3. A general overview, pulling together the reviews and recommendations for all 14 targets, comprises the next section of this report.

Figure 1: NAO's Model for setting validation ratings



Notes

1. "Significant" means that the risk matters; if it does not matter then it is not "significant". Thus a "significant" risk may be mitigated by adequate disclosure. An extremely significant risk (ie "so severe as to undermine the entire data system") cannot be so mitigated.
2. If the risk cannot be cost effectively mitigated and it is so severe as to undermine the entire data system, then adequate disclosure will not compensate; the assessment would then be "red".
3. A disclosure assessment is relevant where the significant uncontrolled risk does not undermine the entire data system and/or there is no reasonable alternative to the data being used despite their limitations.

3 Overview of report conclusions and recommendations

Conclusions

As described above, the data systems underpinning the 14 PSA targets have been thoroughly reviewed by teams comprising Department statisticians and NAO staff. In this section of the report we will summarise the main findings of those reviews (with details appearing later in the report) and organise the emerging recommendations under various broad headings.

The 14 targets have sub-targets, some of which are addressed by different data systems. Targets are split only if there are different data systems and the review findings are significantly different. Therefore, three targets need to be considered in two parts and one in three parts, giving rise to 19 different sets of judgements. Table 1 gives a summary of the validation ratings agreed between the team members for the 14 main targets, while Table 2 gives more detail, split across 19 sub-targets and 9 sets of criteria[2]. For the purpose of Table 2, the following terminology was developed to summarise the often complex judgements made in the reviews in one or two words:

1. **Yes:** The data system fully meets the criteria for this section, with no uncontrolled risks.
2. **Largely:** The data system meets the criteria, but with risks that could be further mitigated. However, the system is essentially fit for purpose.
3. **Not entirely:** There are uncontrolled risks which should be controlled before the system is judged as fit for purpose under these criteria.
4. **No:** The system fails to meet the criteria.
5. **Unsure:** There is insufficient evidence to tell the extent to which the system meets the criteria.

From Table 2 it is clear that data systems judged as green overall are mainly flagged as 'yes' or 'largely' for the individual criteria; 'amber' systems tend to have some 'not entirely' ratings, and those which are deemed to be red contain more than one 'no' rating.

Looking at the different criteria, and counting the number of 'yes' and 'largely' judgements made for each, the largest number is for 1.1 'relevant' and 2.1 'robust' (14) followed by 3.3 'comprehensive' (12) and 1.2 'well-defined' (10). On the other hand, the worst-rated criterion was 2.3 'comparable' (7), followed by 2.2 'reliable' and 2.4 'verifiable' (8), and 3.1 'clear' and 3.2 'transparent' (9).

From this, it might be tempting to make some sweeping generalisations. One could say that the typical data system underpinning the PSA targets was relevant to the target and robust, that to a certain extent reporting was comprehensive, and the system was fairly well-defined. However, it was less likely to be comparable over time, or reliable and verifiable, and reporting was less than clear and transparent.

[2] The nine criteria relate to the three main review headings: 'Specification', 'Operation' and 'Reporting'; the other headings ('Technical Note' and 'Management Controls') are not included in this summary. The former is a specialised area and the latter mainly includes issues covered elsewhere.

Generalisations are invidious, but there is some truth in this one. The data systems in use largely predate the creation of the targets and are in place for a range of purposes, of which tracking progress against the targets is just one. Because of this they tend to be robust, and a majority of them are relevant because they fulfil an organisational need which a target also addresses. Where they are not relevant, this is often because the target was not specified in a way which made tracking progress against it straightforward from existing data systems. This feature of some targets may also mean that they are less well-defined than they might be, and that there have been compromises which tend to lower reliability.

Reporting against targets is variable, with 3.3 'comprehensive' (all aspects included) appearing to be slightly ahead of the other two criteria. In most cases there is scope for improvement here, and this is an area where improvements can be made without vast investment of resources. As part of this, the Technical Note needs to be regularly reviewed and updated in order to fulfil its role relative to the target and the data system.

Continuing the documentation theme, the poor results for 'verifiable' imply that in many cases it is extremely difficult to prove to an outside observer that the systems are error-free and fit for purpose, even if this is in fact the case. The importance of clear and accurate recording of every step in the complex process of deriving target progress measures cannot be overstated. Apart from the audit function, this has a vital role in 'knowledge management' – the maintenance of crucial information for an organisation's business within the organisation rather than just in the mind of a single individual.

Comparability over time is another criterion which overall seems to get poorly rated. This is clearly critical to sensible tracking of progress towards targets, but there is often a conflict between this criterion and improving the quality of the data collected. Where such conflicts appear, it is vital that any non-comparability is well-documented and reported, but also that methods are explored for overcoming it. Two possible strategies are 'back-computation' (calculating what previous years' values would have been under the new system) and 'parallel computation' (continuing to compute results under the old system as well as the new).

Recommendations

General recommendations are classified here under various headings; these will give an indication of the different kinds of work which should be undertaken in order to move all data systems to a point where they are entirely fit for purpose. Detailed recommendations for each target are presented in a later section

1. **Modifying or re-interpreting the target:** In a fair number of cases, there are suggestions that the wording of the target could have been better or the expected improvement quantified. Targets once set cannot be reworded, but greater involvement of analysts in the target setting process may improve the clarity of future targets. See targets 1, 2, 6, 7, 8, 9, 10 and 14.
2. **Studying the effects of risks on target outcomes:** It is often recommended that studies be carried out to assess exactly what the effect of a known risk might be on

a target outcome – if the effect can be quantified it may be possible to ignore the risk or allow for it. See targets 2, 3, 5, 6, 7, 8, 10 and 14.

3. **Technical fixes and validation studies:** There is a similar judgment here, but to use existing data (either already in the system or available on another system) to estimate the effect of the risk. See targets 2, 3, 6, 7, 8, 9, 12, 13 and 14.
4. **Moving from sample to census:** This kind of recommendation is designed to improve the accuracy of estimation, sometimes by using existing administrative data. For instances where alternatives are not available, there needs to be clear guidance on the use and interpretation of sample based estimates and the associated confidence intervals. See targets 1 and 4.
5. **Improved training or guidance:** This is suggested in order to improve the quality of the data collected by ensuring that training or guidelines are much clearer about what is required from those providing the data. See targets 1 and 8.
6. **Improving measurements or systems:** In a few cases reviewers recommended fundamental changes to systems, or the development of new measures in order to overcome the perceived risks with the existing system. See targets 7, 11, 12, 13 and 14.
7. **Improved/fuller reporting and disclosure:** This was a common theme throughout, with better and more open reporting to include known risks and possible sources of error and uncertainty in the reporting of progress towards targets. See targets 1, 2, 3, 6, 7, 8, 9, 10, 11, 12, 13, 14.
8. **Consistency of reporting:** Alongside the above, there were also calls for consistency across different reports from different departments, or even within the Department itself. See targets 10 and 13.
9. **Parallel reporting for comparability:** In some cases there were recommendations that parallel calculations should be carried out and reported to ensure comparability of outcomes when assessing change over time. See targets 3, 4, 6, 7, 8 and 13.
10. **Enhanced quality assurance:** There were several calls for QA processes to be improved, including regular reviews and audits of key elements of the data system which might give rise to significant risks. See targets 4, 5, 6, 7, 8, 9, and 11.
11. **Better documentation:** This was a fairly common theme, linked to the concept of verifiability of data systems. See targets 2, 3, 5, 10, 12, 13 and 14.

From the above, it is clear that there is a wide range of recommended improvements which could be implemented, with a focus on enhancing the data systems with regard to their role in measuring progress towards PSA targets. Not all of these recommendations will be approved or implemented, although many (such as those related to documentation and reporting) are relatively low-cost and do not require any radical changes to existing systems.

It should also be remembered that, although these data systems do underpin the PSA targets, they all have other purposes and functions, both for the Department and for a wider community of users. Although the focus here has been on the target-reporting function, suggested enhancements may well bring wider benefits throughout the Department's remit.

One of the key recommendations that should be carried forward by all involved in PSA Targets is that alluded to in the first point in the above list. It is vitally important that in future analysts and those who understand data systems are involved in the

development of targets, to ensure that they are in fact measurable and that appropriate data systems do exist to underpin them.

It has been suggested that all data systems underpinning PSA targets should be National Statistics, as this status gives a guarantee of some minimum quality standards. Reference to the appropriate part of the Office for National Statistics (ONS) website[3] does indeed set out a set of principles for National Statistics, including relevance, integrity, quality, accessibility etc. Further investigation of the list of current National Statistics seems to indicate that all but four of the data systems reviewed in this report are in fact National Statistics (or there seem to exist National Statistics which are closely related to them). The exceptions are those for the third element of Target 2 (Childcare Approval Scheme (CAS) data), the Abortion Notifications for Target 3 and also Target 9 (both the Annual PE, School Sport & Club Links (PESSCL) school survey and the Ofsted reports).

In many ways this is good news, in that most of the relevant data systems are signed up to the standards implicit in National Statistics. However, the fact that our review has still shown only two 'green' validation ratings implies that being a National Statistic does not, in itself, guarantee total fitness of purpose in terms of measuring progress to PSA targets. It would seem that becoming a National Statistic should be a minimum requirement for all such data systems, but it should be regarded as a necessary but not sufficient condition for the award of a 'green' rating.

[3] http://www.statistics.gov.uk/about/ns_ons/nsproducts/default.asp

Table 1: Current Provisional Validation Ratings for PSA Targets*

SR2004 Target	Brief description	Joint target with	Systems	Validation rating
1	Improve children's communication, social and emotional development	DWP	Foundation Stage Profile (FSP)	Amber
2	No-one working Households	DWP	Ofsted data on registered childcare, Household below average income (HBAI) from DWP, & CAS data	Green/Amber/ White
3	Under 18 conception rate	Dept. of Health (DH)	ONS conception data	Green
4	Obesity	DH & DCMS	Health Survey of England (HSE)	White
5	Looked after children		Social Services Department SSDA903/Outcomes Collection 2 (OC2) data collection	Amber
6	Maths and English standards for 11 year olds		Results of national curriculum Key stage 2 tests	Amber
7	Maths and English standards for 14 year olds		Results of national curriculum key stage 3 test	Amber
8	School attendance		Registration data from schools	Amber
9	Sporting opportunities	DCMS	PESSCL Survey, Ofsted reports	Amber
10	5 GCSE for 16 year olds		School Examination Results Analysis Project (SERAP), Pupil Level Annual Schools Census (PLASC)	Amber
11	Level 2 for 19 year olds		PLASC, SERAP, National Information System for Vocational Qualifications (NISVQ), Individual Learner Record (ILR)	Amber
12	Young people not in education, employment or training (NEET)		ONS/Government Actuary's Dept. (GAD) population estimates, school census (PLASC), ILR, HESA student record, Labour Force Survey (LFS)	Amber
13	Skills for life		ILR, Offender Learning & Skills Unit (OLSU) data, Jobcentre+ data, LFS	Red/ Amber
14	Higher education		Higher Education Initial Participation Rate (HEIPR) from Higher Education Statistics Agency (HESA) data & ILR, HESA (Fair Access, Rates of non-completion)	Amber/ Red / Red

* The ratings shown are those agreed between the members of the two-person team reviewing each data system (DfES statistician and NAO examiner).

Table 2: Summaries of data systems by review criteria

			Specification		Operation				Reporting		
Target	**Data systems**	**Validation rating***	**1.1 Relevant****	**1.2 Well-defined**	**2.1 Robust**	**2.2 Reliable**	**2.3 Comparable**	**2.4 Verifiable**	**3.1 Clear**	**3.2 Trans-parent**	**3.3 Compre-hensive**
1	FSP	Amber	Largely	Not entirely	Largely	Largely	Not entirely	Largely	Unsure	Unsure	Unsure
2 (i)	Childcare stock (Ofsted)	Green	Yes	Yes	Largely	Largely	Largely	Largely	Largely	Largely	Largely
2 (ii)	HBAI	Amber	Yes	Largely	Not entirely	Not entirely	Not entirely	Not entirely	Not entirely	Yes	Largely
2 (iii)	CAS data	White	Yes	Not entirely	Unsure	Unsure	Unsure	Unsure	Not entirely	Not entirely	Largely
3	ONS conception data	Green	Yes	Yes	Largely	Largely	Yes	Not entirely	Yes	Yes	Largely
4	HSE	White	Largely	Largely	Largely	Largely	Not entirely	Unsure	Unsure	Unsure	Unsure
5	SSDA903 and OC2	Amber	Yes	Not entirely	Largely	Yes	Largely	Not entirely	Largely	Not entirely	Not entirely
6	KS2 assessment data	Amber	Largely	Largely	Yes	Not entirely	Not entirely	Largely	Largely	Largely	Largely
7 (i), (ii), (iv) & (v)	KS3 assessment data	Amber	Largely	Largely	Yes	Not entirely	Not entirely	Largely	Largely	Largely	Largely
7 (iii)	KS3 ICT	Amber	Not entirely	Not entirely	Yes	Not entirely	Not entirely	Not entirely	Largely	Largely	Largely
8	Registration data from LEAs	Amber	Not entirely	Not entirely	Not entirely	Not entirely	Not entirely	Not entirely	Not entirely	Largely	Largely
9	PESSCL & Ofsted reports	Amber	Largely	No	Largely	Not entirely	Not entirely	Not entirely	Not entirely	Not entirely	Not entirely
10	SERAP, PLASC	Amber	Yes	Largely	Largely	Yes	Not entirely	Largely	Largely	Not entirely	Yes
11	PLASC; SERAP; NISVQ; ILR	Amber	Yes	Largely	Yes	Yes	Largely	Not entirely	Largely	Yes	Yes
12	ONS/GAD, PLASC, ILR, HESA, LFS	Amber	Not entirely	Largely	Largely	Not entirely	Not entirely	Largely	Not entirely	Not entirely	Not entirely

Target	Data systems	Validation rating	1.1 Relevant	1.2 Well-defined	2.1 Robust	2.2 Reliable	2.3 Comparable	2.4 Verifiable	3.1 Clear	3.2 Trans-parent	3.3 Compre-hensive
13(i)	ILR, OLSU, Jobcentre+	Red	No	No	Largely	Not entirely	Largely	Largely	No	No	No
13 (ii)	LFS	Amber	Largely	No	Not entirely	Not entirely	Largely	No	No	Not entirely	Largely
14(i)	HESA, ILR	Amber	Largely	Largely	Largely	Largely	Largely	Largely	Largely	Largely	Largely
14 (ii) & (iii)	HESA	Red	No	No	Unsure	Unsure	Unsure	Unsure	Unsure	Unsure	Unsure

***** Validation ratings: White: Too early to form a view on the data system; Green: Effective controls have operated and disclosure is adequate; Amber: Data system and/or reporting need strengthening; Red: Significant weaknesses exist.

****** Judgements: Yes: The data system fully meets the criteria for this section, with no uncontrolled risks;
Largely: The data system meets the criteria, but with risks that could be further mitigated. However, the system is essentially fit for purpose;
Not entirely: There are uncontrolled risks which should be controlled before the system is judged as fit for purpose under these criteria;
No: The system fails to meet the criteria;
Unsure: There is insufficient evidence to tell the extent to which the system meets the criteria.

4 Summaries of reviews and recommendations for each target

In this section of the report we will provide brief summaries of the reviews of each target's data system(s).

PSA Target 1

Improve children's communication, social and emotional development so that by 2008 50%* of children reach a good level of development at the end of the Foundation Stage and reduce inequalities between the level of development achieved by children in the 20% most disadvantaged areas and the rest of England. (Joint with DWP)

* - provisional target.

Data System:

FSP data collection.

Summary of review

The two elements of the target (improvement in communication, social and emotional development and reduction in inequalities) are collected in two different formats. The improvement in children's communication, social and emotional development is measured through a 10% sample of children's individual records which is taken in the July following the child's fifth birthday. These records show whether a child achieved a 'good level' (6 points or more) on all 7 assessment scales (Personal, Social & Emotional Development (PSE) has 3 scales and Communication, Language and Literature (CLL) has 4 scales). The reduction in inequalities will be measured via the Local Authority (LA) Super Output Area[4] (SOA) aggregate return. The department provides each LA with a list of SOA identifiers and postcodes to enable them to complete the aggregate return.

The system is therefore largely relevant to the target, although there remain some inconsistencies in the assessment and moderation of the pupil measures, which mean it is not entirely well-defined. The use of a sample of records is a source of potential bias. There are extensive quality assurance procedures in place at different stages of data collection, which means that the system is largely robust and reliable as well as verifiable, although the consequences of the 10% sample may need further documentation. A software change for 2006 should enable progress to the target to be measured from the aggregate data rather than the 10% sample, although checks will be required to ensure consistency with the sample results. The change from most disadvantaged 20% of postcodes to most disadvantaged 30% of Super Output Areas (SOAs) may pose a threat to comparability.

[4] Super Output Areas are a new geography for the collection and publication of small area statistics. The SOA layers form a hierarchy based on aggregations of Output Areas (OAs). They are better for statistical comparison as they are of much more consistent size and each layer has a specified minimum population to avoid the risk of data disclosure (releasing data that could allow the identification of individuals). SOAs are not subject to frequent boundary change, so are more suitable for comparison over time. There are 32,482 Lower Layer SOAs in England.

There has been no reporting as yet on the target, so it is not yet possible to judge if this is clear, transparent and comprehensive. The Technical Note needs to be updated to reflect the change to SOAs mentioned above.

Target 1 Recommendations

- At present the 10% sample appears to be robust enough to provide a national evaluation of progress against the target. However, it is recommended that this strategy be reconsidered, as it may be important to look at progress at a more disaggregated level, and the LAs already collect 100% of individual child level data. The change to the 2006 data collection may be enough to satisfy this.
- LAs should be encouraged to improve communications with private, voluntary and independent settings (PVIs) to reduce non-response to the 100% data collection.
- The recommendation by Oxford University to use IDACI (Income Deprivation Affecting Children Index) instead of the present IMD2004 should be given serious consideration.
- We recommend quantifying the desired reduction to make the target more focused and challenging.
- LAs should be encouraged to develop systems for ensuring that all practitioners are effectively trained in the statutory curriculum and its assessment and to develop a moderation plan. Information on the moderation of results and the training of teachers should continue to be included in the target reporting.
- Inconsistencies between data collected in 2005 and future years should be fully reported in any publications reporting progress against the target.
- Key findings at the LA level could be included in the Statistical First release (SFR) to make the data collection appear more relevant to LAs.
- The Technical Note needs to be updated to reflect the change to measuring inequalities based on 30% most deprived SOAs.

PSA Target 2

As a contribution to reducing the proportion of children living in households where no one is working, by 2008

i. increase the stock of Ofsted-registered childcare by 10%
ii. increase the take-up of formal childcare by low income working families by 50%; and
iii. introduce by April 2005, a successful light-touch childcare approval scheme.

(Joint with DWP)

Data Systems:

(i) Stock of Childcare (Ofsted)
(ii) Take up by lower working families – Households below average income (HBAI - DWP)
(iii) Light Touch Approval Scheme (Contractor)

Summary of review

The data systems are all relevant to the target, and well-defined for part (i). Part (ii) is largely well-defined, but part (iii) is not fully defined.

For part (i), the system is largely robust and reliable, with just 3.4% of providers having missing data on number of places which needs to be imputed. For (ii), the data is collected via the Family Resources Survey (FRS) which has a 64% response rate. Although the HBAI is weighted to adjust for non-response bias, there is no estimate of sampling uncertainty, so the data system has to be rated as not entirely robust and reliable. As the data system for part (iii) is not yet fully defined, its operational characteristics cannot be determined.

For both parts (i) and (ii) there are risks to comparability. For (i) the number of null providers is expected to decrease over time and hence if the imputation procedure is inaccurate then the figures may no longer be comparable. For (ii) there has been a change to the childcare question in the survey. Although system (i) is well-documented and hence verifiable, the documentation for (ii) is less complete, posing risks to its verifiability.

The reporting of system (i) is generally clear, transparent and comprehensive, but there are elements which could be improved (e.g. reporting of imputation). The reporting for part (ii) is not entirely clear (definition of lower income households not given in the Autumn Performance Report (APR)), although it is transparent and comprehensive. There are also problems with the clearness and transparency of the reporting for part (iii).

The Technical Note is currently under review - the section on the CAS will need to be updated to reflect the future following any changes to the indicators and the potential impact of the Childcare Bill on the Ofsted childcare register and the CAS.

Target 2 Recommendations

The Department should work with Ofsted and DWP on the following:

General

- The DfES APR should be re-formatted to make it clearer that the whole target is joint with DWP and not just the third bullet.
- The DWP and DfES should share their draft reports for this target before publication to ensure consistency of reporting.

Part (i)

- An analysis should be prepared showing the effect that null providers might have on the target, both currently and in the future.
- Documentation should explain clearly the imputation procedure that is used for null providers.
- The Department should clarify that the target refers to the maximum capacity of childcare as judged by Ofsted inspectors not the actual number of places filled, referring to National Standards for childcare in the Technical Note to explain how the maximum capacity is determined.
- Additional disclosure relating to estimated places on the Ofsted childcare register should be provided.

Part (ii)

- An assessment should be made of the role that sampling variation could play in measuring the childcare take-up in lower income households.
- An approximate estimate of the impact of the change in FRS interview methodology should be made by comparing LFS to FRS childcare take-up results for all households.

Part (iii)

- The Department should agree a final version of part iii) of the target with Treasury.
- Disclosures relating to the Childcare Approval System and to explain the impact of the childcare reforms should be improved.

PSA Target 3

Reduce the under-18 conception rate by 50% by 2010 as part of a broader strategy to improve sexual health. (Joint with DH)

Data Systems:

- Birth Registrations (ONS)
- Abortion notifications (DH)
- Mid-year population estimates (ONS)

Summary of review

In general the system is relevant and well-defined. Data is derived for the target population from all birth registrations and abortion notifications in each year and conception statistics are released around 14 months after the end of the year to which they relate. Overall, due to adequate controls in place, the system is robust and reliable. For live births, the date of conception is calculated assuming a gestation period of 38 weeks. Age of the woman at conception is also derived using gestation. Some births are premature and this could have an impact on the accuracy of the conception statistics. At present the impact is not known, although the same methodology has been used consistently over time and is fully reported in the notes to the published statistics; therefore the system may be regarded as comparable.

Due to a lack of adequate documentation in some areas the system is not entirely verifiable. The reporting of the data appears to be clear and transparent, and largely comprehensive – the main risk is the 14-month delay in the statistics being available. The Technical Note briefly mentions how the target is derived but does not give an example or indicate what the base rate is or the current position. The data systems are long-established and two out of three are National Statistics; management controls appear to be adequate for this target.

Target 3 Recommendations

The Department should work with ONS and DH on the following:

- The ONS are hoping to obtain the actual gestation period for live births through data collected by the NHS Connecting for Health. Work will need to be done to estimate the impact this will have on the birth/conception rate and if this is significant then this data needs to be published as the true birth/conception rate. However, the conception rate based on the assumed gestation period will still have to be produced to ensure consistent monitoring of the target to 2010
- It is important that documentation on all quality assurance processes (both in the capture of the data and production of the statistics) should be readily available, especially with production of the conception statistics soon to transfer to another team. It is recommended that documentation be produced as soon as possible.
- Although reporting towards the target would not be affected, it is recommended that efforts be made to obtain sources or estimates of miscarriage data that could be combined with the current data to improve the accuracy of conception rates.
- Disclosure regarding the limitations of the data could be improved (i.e. miscarriages).
- The Technical Note should include an outline of the target population, a brief description of each data source, an indication of how the rate is calculated with respect to the latest data available, links to current documentation and contact details of where to go for further information.
- Plans for registrars to submit birth registrations online in 2006 should be monitored, to determine if this improves the quality and timeliness of the data. Similarly, the pilot web-based submission system for abortion notifications should be monitored.
- It is recommended that ways of producing more timely estimates of the conception data be investigated.

PSA Target 4

Halt the year-on-year rise in obesity among children under 11 by 2010 in the context of a broader strategy to tackle obesity in the population as whole. (Joint with Dept. for Culture, Media and Sport (DCMS) and DH)

Data Systems:

The annual 'Health Survey for England' (HSE), owned by The Information Centre for Health and Social Care (IC), and carried out by the Joint Health Surveys Unit (JHSU) of NatCen (the National Centre for Social Research) and University College London (UCL, Department of Epidemiology and Public Health. In order to enable reporting on the target, JHSU will analyse the HSE data to isolate the data specific to children aged 2 to 10.

Summary of review

A child boost is to be introduced in the final 5 months of 2005 and for every full year after this, so that approximately 12,000 children aged two to ten will be included in each 3 year average. This measure has been put into place specifically to aid the measurement of the PSA Target. The sample is well-defined and relevant to the population, and the boosting will add to the validity of the data. Confidence intervals based on the sample size have been published, and should be regularly updated for each successive survey.

The sample is a multi-stage stratified probability sampling design and the sampling frame is the small user postcode address file. The system is largely robust and reliable, but non-response weighting was introduced in 2003. This will affect the comparability of the baseline data 2002/03/04 as 2002 data is not weighted.

The HSE is a well established system. Clear documentation about its methodology and results are publicly available. Further work is however required: though JHSU has recently isolated the data relevant to children aged 2-10 from the wider data to date the analysis has not been carried out, and the system is not entirely verifiable. There has been no reporting to date against the target, so it is not possible to assess this element of the system.

The Technical Note is generally clear and concise, and all the specifications are well defined. It would benefit from the inclusion of a link to the Health Survey to give further information on the methodology for the HSE. Controls have not yet been established to ensure JHSU's analysis of the HSE for the purposes of reporting on the target is reliable.

Target 4 Recommendations

The Department should work with DH on the following:

- The possibility of validating the HSE results for school-age children using PLASC data should be investigated.
- Confidence intervals taking account of the sample properties should continue to be published alongside the reported estimates.
- Robust procedure should be devised to ensure that the analysis undertaken by JHSU is reliable.
- If Body Mass Index (BMI) measures change for any aspect of the target this should be well documented, and if necessary old measures also computed to ensure comparability in measuring the target.
- The Technical Note should be updated to include all relevant information, including the child boost, weighting in households with over 2 children, a link to information on the HSE. There are also errors and unclear statements which need to be amended – for example the Target year is given as 2010, but would be better stated as 2008/09/10; the '2010-11 financial year' should be '2011 calendar year'.

PSA Target 5

Narrow the gap in educational achievement between looked after children and that of their peers, and improve their educational support and the stability of their lives so that by 2008, 80% of children under 16 who have been looked after for 2.5 or more years will have been living in the same placement for at least 2 years, or are placed for adoption.

Data Systems:

Social Services Departments Annual 903 Collection (SSDA903) and OC2.

Summary of review

The data system is relevant but not entirely well-defined; the wording of the target can lead to misinterpretation of exactly what is to be measured. The system is robust and reliable, with a small risk of non-response bias. Controls appear to be comprehensive on the whole, and the only inadequacy appears to be the lack of auditing done to ensure that the controls are working.

There will be a change in the collection regarding Special Guardianship, which is a new reason for leaving care. This may reduce the numbers of children in care for long periods. This change is well recognised (in the Technical Note) and the data is being collected in 2006 so that the extent of any change can be assessed and prepared for. Otherwise, the data system is comparable. The data appears to be collected and transferred in a secure environment, but the system is not entirely verifiable, due to a lack of documentation on data capture at the LA level and data analysis.

The reporting under the system is not entirely clear, transparent and comprehensive, as the reporting to Parliament only gives partial information. The Technical Note is largely comprehensive and clear about the stability component of the target, but is lacking in other areas.

Target 5 Recommendations

- The Technical Note should clarify whether improving educational support is intended to be a target, and if so a data system for measuring this target needs to be defined
- The Technical Note should be amended in various ways, to increase clarity and remove inaccuracies.
- The Technical Note should be amended to define a SMART-compliant target for narrowing the gap in educational achievement.
- Analysis should be carried out to assess the extent to which changes (e.g. special guardianships) are likely to affect the statistics related to the target.
- Official audits of the data processing systems should be conducted.
- FAQs and Guidance notes should be reviewed and updated to ensure inclusion of all relevant, recurring queries.
- Desk instructions detailing the internal systems and processes should be written.
- Direct checks should be instituted by the Department to ensure records are completed for all looked after children.
- The Department should tap into audits done by the Commission for Social Care Inspection (CSCI), the Audit Commission (AC) and Ofsted.
- The Department should ensure its documentation on web data transfer and security is up to date and accessible, and a certificate on the security of the web system should be obtained.
- The Department should put in place a memorandum of understanding with the Audit Commission, in order to ensure that the findings of audit work with respect to local authority control environment and performance indicators can be assessed for their impact on the PSA target.
- The reporting against the target needs to be clarified on a number of issues, and should cover all aspects of the target.

PSA Target 6

Raise standards in English and maths so that:

i. by 2006, 85% of 11-year-olds achieve level 4 or above, with this level of performance sustained to 2008; and
ii. by 2008, the proportion of schools in which fewer than 65% of pupils achieve level 4 or above is reduced by 40%.

Data Systems:

The results of pupils in statutory Key Stage 2 (KS2) assessments in English and mathematics usually taken at age 11. Results in English and mathematics are confirmed by schools as part of the Achievement and Attainment Tables (AAT) checking exercise.

Summary of review

The data system is relevant and covers all aspects of the target. However, the coverage of the data does not match exactly to the definition of the target due to the exclusion of independent schools (unless they 'opt in') and pupils taking assessments out of year. Analysis indicates that the exclusion of such pupils does not have a material impact on the statistics. Other risks relate to moderation of test-taking, which seem to be well-controlled, and the data system is judged to be largely relevant and well-defined relative to the target.

The system is largely robust, in that the statistics produced reflect the results obtained by the population despite a small number of omissions and exclusions. Risks related to the reliability of the measures used (i.e. KS2 tests) are regarded by the Qualifications and Curriculum Agency (QCA) and other stakeholders as generally well-controlled, although there remain concerns and a number of comments have been made regarding risks to reliability due to factors such as borderlining. Moreover, issues have been raised about the extent to which assessment results are comparable over time and there is a body of documented work and debate on this topic. The Statistics Commission concluded that the system was largely fit for purpose, and due to the well-documented nature of most of the system it appears to be by and large verifiable. The development of a new system to improve reliability and comparability of national performance measures would be possible, but would also be expensive, and there is ongoing debate about this would be a cost-effective solution.

The system for reporting is largely clear, transparent and comprehensive. In some cases the reporting does not spell out the exact population tested, and results could be reported to one decimal place to advantage. Some small amendments are needed to the Technical Note to ensure it is completely accurate.

Target 6[5] Recommendations

The Department should work with QCA, Data Collection Agency (DCA) and National Assessment Agency (NAA) on the following:

- Work to validate the maintenance of standards over time should be ongoing.
- The effects of non-inclusion of independent schools and the inclusion of results for pupils not in the specified age range should be monitored.
- Target figures should be reported to one decimal place to improve the accuracy of the information and reduce the apparent volatility of reported results.
- The exclusion of small schools and overseas pupils from school level targets should continue to be monitored, and disclosure on the number of small schools excluded from the target in the Statistical First Return should be improved.
- Further monitoring of the effects of borderlining and the upward shift due to schools requesting reviews should be made and the impact on the target figures quantified, including investigation of possibly shifting the borderlining zone.
- The planned reviews to improve the tracking and reporting of errors between the (DCA), (NAA) and the Department should have a high priority.
- A change to marking to ensure that markers only have to enter the correct result on the mark sheet and do no calculations should be seriously considered and piloted, to determine if it would reduce errors and borderlining effects.
- Further monitoring of whistle-blowing allegations should be carried out.
- Disclosure on independent schools in SFR, departmental report and PSA target report should be improved, and consideration given to re-specifying the coverage of the target in the 2007 spending review to exclude independent schools.
- The Technical Notes should be clarified in various ways, including end of KS2 not age 11 publication, overseas pupils, and bias in the system caused by amendments being made only in a pupil's favour.

[5] These comments for Target 6 (KS2) also apply to Target 7 (KS3).

PSA Target 7

Raise standards in English, maths, ICT and science in secondary education so that:

i. by 2007, 85% of 14-year-olds achieve level 5 or above in English, maths and ICT (80% in science) nationally, with this level of performance sustained to 2008; and
ii. by 2008, in all schools at least 50% of pupils achieve level 5 or above in each of English, maths and science.

Data Systems:

The results of pupils in statutory Key Stage 3 assessments in English, mathematics, science and ICT usually taken at age 14. Results in English, mathematics and science are confirmed by schools as part of the Achievement and Attainment Tables checking exercise.

Summary of review

This data system is almost exactly parallel to that for Target 6 above, with the exception of ICT assessment (currently measured using teacher assessments). For that reason, comments for Target 6 should be taken as applying to this target also, and this summary will focus on the review of the ICT component.

QCA is currently piloting online tests for ICT with a view to introducing a full KS3 test in 2007/08. For the time being, reporting of ICT results continues to rely on teacher assessments. These assessments are not moderated between teachers, schools or years, and therefore this element of the system is not well-defined even if it is relevant to the target.

Although there is a relatively robust system for collecting ICT assessments, this is not reliable, comparable or verifiable. The reporting of the system is largely clear, transparent and comprehensive. The Technical Note explains that ICT is based on teacher assessment, but does not discuss the consequent issues.

Another issue which affects this target is the wording of the last element, with its reference to 'all schools'. The Statistics Commission has criticised this form of target, but this does not directly affect the data system.

Target 7 Recommendations

The recommendations for Target 6 also apply to this target; additional recommendations given below are specific to this target.

- Development of a valid ICT assessment should be given an extremely high priority.
- All references to the ICT component of the target should stress the unreliability of the current measure and the risks to comparability over time.

PSA Target 8

Improve levels of school attendance so that by 2008, school absence is reduced by 8% compared to 2003.

Data Systems:

- Statutory School Registration
- DfES School Performance Information: Absence in Schools
- SFR 49 (Pupil Absence in Schools in England)
- 2005/06 onwards: PLASC (Pupil Level Annual School Census) – but note this will be replaced by the School Census from 2006.

Summary of review

The most significant threat to the relevance of the data system is the fact that post-registration truancy is not captured. There are also problems with the consistency and accuracy of data collection from a significant number of schools, and ongoing issues with definitions of authorised and unauthorised absence, and with children not appearing on school rolls for various reasons, mean that the system is not entirely well-defined. The wording of the target is also open to different interpretations.

There are quite complicated risk considerations for robustness due to the nature of the data collection at school level, and equally complex controls exercised by the Department through legislation and guidance. For these reasons, the robustness of the data system is subject to question. Response rates for the absence data collection are high, but it is clear from the small incidence of incomplete or non-response that the statutory obligation to provide school absence returns does not operate perfectly. It is not known how effective the published guidance on the use of the Approved Educational Activity category is with relation to the data collection, but there is some evidence to suggest that it is insufficient to mitigate the risk of inaccurate data recording, and therefore the system is not entirely reliable.

There may be threats to comparability in current attempts to improve data quality, especially with the move to termly data collection at the pupil level. The current system is not entirely verifiable, mainly because no controls are in place to ensure the validity of the data collection at school level.

The reporting is not entirely clear, mainly due to an issue with the HM Treasury website. It is largely transparent, with a couple of minor issues that should be included in reporting. The breakdown of the national absence rate into authorised and unauthorised absence is published regularly as a National Statistic. Although this is no longer relevant to the target itself, it does make an important contribution to comprehensive reporting of the data. The Technical Note is up to date as far as it goes, but would benefit from expansion.

Target 8 Recommendations

- Independent research should be conducted, via procurement of a sample of pupil registration data and school aggregation procedures, which could provide a much needed indication of the likely scale of underreporting of pupil absence.
- The Department should consider the effect of independent schools on the data series and whether they should be removed from the target.
- The Department should conduct estimates of the effect of including final half-term data on the national target rate.
- Once pupil-level data is established, a system should be introduced to monitor separately the absence rate of subsequently excluded pupils, to check for any bias resulting from disproportionate levels of exclusion among persistent truants.
- A concerted publicity and training effort should be undertaken when the new guidance is published, to maximise the improvement of data quality.
- Publication of the total number of sessions alongside the percentage absence figures, would enable assessment of whether variation in the denominator has been a factor in any future trends in absence.
- Consideration should be given to requiring schools to submit an electronic copy of their raw registration data for checking in certain cases.
- The target information on the HMT website should be updated.
- The SFR should note that the overall absence figures used to measure the PSA target include absence rates for independent schools.
- The Technical Note should be substantially updated, to include descriptions of all identified risks and documentation of all relevant systems and procedures.
- When PLASC is replaced by the School Census in 2006 and 2007, an assessment of the effect of the change should be reported.

PSA Target 9

Enhance the take-up of sporting opportunities by 5 to 16 year olds so that the percentage of school children in England who spend a minimum of two hours each week on high quality PE and school sport within and beyond the curriculum increases from 25% in 2002 to 75% by 2006 and to 85% by 2008, and to at least 75% in each School Sport Partnership by 2008. (Joint with DCMS)

Data Systems:

PESSCL School Survey; Ofsted subject reports

Summary of review

The annual PESSCL survey is used to measure the quantity of provision, but is only completed by schools in a School Sports Partnership on behalf of pupils. The proportion of English maintained schools in a Partnership is increasing. As plans to collect this data through PLASC were rejected, we believe there is no better way of collecting the data on the quantity of provision, and the system is basically relevant. Lack of details in Ofsted's figures on quality lead to a risk that the system is poorly defined.

The system is largely robust and reliable with regard to quantity of provision but less so for quality. There is a logical discontinuity between the definition of quality in the target (defined for pupils) and information from Ofsted inspections (defined for lessons) which renders this aspect of the data system not entirely robust and reliable.

The size of the survey changes over time as more schools join Partnerships, which means the system is not entirely comparable. The PESSCL survey is now a well established system, and data on quantity of provision is largely verifiable; however, this is not true of the use of Ofsted data to define quality.

Reporting does not cover all aspects of the target and is not clear about the limitations of the data. Furthermore, the baseline gives a misleading impression of progress. For these reasons the reporting is not entirely clear, transparent and comprehensive. Although the Technical Note is up to date, it should be strengthened by the disclosure of fuller information on the data systems and the underlying assumptions.

Target 9 Recommendations

- Greater use should be made of the Youth Sport Trust (YST) reports to validate the survey responses prior to publication of the results.
- We recommend that DfES should ensure that it understands and documents the methodology used to calculate this data, and takes steps to satisfy itself that Ofsted's systems are likely to generate relevant, reliable and comparable information.
- DfES should ensure that future targets are specific and measurable (e.g. as regards measuring quality) before they are agreed with Treasury.
- The number of schools in Partnerships should be reported alongside results.
- Disclosures need to be strengthened to report a number of factors, including that the data cannot currently give figures for all English maintained schools and how the level of provision in 2006 was calculated.
- The Technical Note should be amended to give greater disclosure on a range of relevant issues.

PSA Target 10

By 2008, 60% of those aged 16 to achieve the equivalent of 5 GCSEs at grades A*-C; and in all schools, at least 20% of pupils to achieve this standard by 2004, rising to 25% by 2006 and 30% by 2008.

Data Systems:

SERAP, PLASC

Summary of review

The system is essentially a 'census' of all qualifications gained by 15 year olds in England each year and is thus relevant to the target. It is also well-defined, although the increasing number of vocational qualifications included may be a challenge to this.

The system is largely robust, although there is a risk that schools may tend only to challenge results they feel are too low, which may result in a slight upward bias. The extensive number of validation checks built into the system assure that it is reliable. There is a risk that improvements over time are due to new qualifications or better data collection rather than genuine progress, which may mean that the system is not entirely comparable. Quality assurance and other detailed checks mean that the system is largely verifiable, though some documentation could be improved.

Statistics are reported in Departmental publications in accordance with National Statistics principles and protocols, which ensures exceptions and problems are clearly reported (such as data collection issues). However, there are some inconsistencies in the reporting of the target between the SFR and departmental reports and the Technical Note, which means that the system is not transparent between publications.

The Technical Note appears to pre-date the decision to start collecting extra vocational equivalents to monitor the target, and needs to be updated.

Target 10 Recommendations

The Department should work with QCA on the following:

- The Department should conduct some research to assess the effect of including vocational qualifications in the data system.
- The Department should investigate the scale and impact of revisions by schools.
- The Department should improve reporting of the target measurement to clarify (and quantify) that any performance improvements may in part be due to new qualifications or better data collection.
- The current standards monitoring should be developed by making the results and methodology more public and frequent, and including more comparisons of GCSE results against other sources of attainment information (Program for International Student Assessment etc).
- More standard documentation should be produced and readily available.
- For clarity we recommend that the PSA target be reworded in line with the SFR.
- Departmental Reports should show progress against the targets in the same way as the SFR.
- Consideration should be given to looking at ways of 'slimming-down' the main SFR document in order to improve clarity and accessibility.
- The Technical Note should be updated in a number of ways to improve clarity and completeness.

PSA Target 11

Increase the proportion of 19-year-olds who achieve at least level 2 by 3 percentage points between 2004 and 2006, and a further 2 percentage points between 2006 and 2008, and increase the proportion of young people who achieve level 3.

Data Systems:

PLASC; SERAP; National Information System for Vocational Qualifications (NISVQ); ILR.

Summary of review

The system is relevant and well-defined as it combines data from several databases which span the population. The main risk is that new qualifications that have no previous equivalent are being introduced each year. Comparability of standards over time is also at risk because qualifications are largely based on new examinations each year. The system is robust and reliable, as the overall process was reviewed and the subsequent report signed off by the Head of Profession for Statistics and the National Statistician. There is a matching rate of about 90% between datasets, but a study carried out estimates that matching error may result in a +/-0.1% error in the final target figure reported.

Apart from issues about new qualifications and maintenance of standards mentioned above, the system is largely comparable. However, the system is not sufficiently verifiable. There is a need for the procedures to be more formalised, with collection and validation to be more transparently recorded.

The outcome is a National Statistic, and the reporting is largely clear, transparent and comprehensive, with a few gaps in reporting need to be amended. The Technical Note is not up to date and also needs to be amended in some respects.

Target 11 Recommendations

- The Department should implement a more robust project management processes, including a clearer governance structure with defined roles and responsibilities.
- Formal procedures for calculating the target should be produced, including the templates for recording the outputs of the testing and validation processes.
- Reports should disclose that some proportion of the target age group will have reached level 2 through qualifications that did not exist when the baseline was set, as well as the change of the data collection process for the 2005/2006 period, and its potential implication for error.
- The negotiation for a single data matching contractor should be expedited.
- The Department should consider obtaining the necessary funding to cover other establishments (for pupils not educated in schools, but in hospitals or at home, for example) or imposing stricter guidelines on LAs to provide their data.
- The Technical Note should be updated with some clarifications and corrections.

PSA Target 12

Reduce the proportion of young people not in education, employment or training (NEET) by 2 percentage points by 2010.

Data Systems:

ONS/GAD population estimates;
DfES Schools census data (PLASC & SLASC);
Learning & Skills Council (LSC) Individualised Learner Record (ILR);
HESA Student Record;
ONS Labour Force Survey (LFS);

Summary of review

The population relevant to the target is the population of 16-18 year olds in England, who are not in education, employment or training. This information is not obtained directly, but is determined by a process of elimination, and therefore the data system is not entirely relevant (although it is not clear if there is any viable alternative to access the required population). Although the system is well-defined, in terms of the processes used to derive the key statistic, there are a number of significant risks which are not well-controlled. These include discrepancies in the population of 14 year-olds, and the omission of students studying outside England, those undertaking personal development opportunities or custodial sentences, and refugees.

The system is largely robust, and errors due to LFS sampling are estimated and presented. It is not clear what the actual reliability of the estimate may be, in the light of significant risks due to non-response bias and proxy responses in the LFS. Sampling error estimates are published, but may not be the major source of unreliability. Although the method for calculating the statistic has remained stable, the underlying data sources have undergone changes which may affect comparability. Most data systems underpinning this calculation are well-documented and verifiable – the main exception is LFS.

Neither the SFR nor the Technical Note state clearly the inclusions in the NEET group after education, employment and training are considered, namely: gap years, personal development, custodial sentences and refugees ineligible for work. This information must be assumed on the part of the reader. For these and other reasons the reporting is not entirely clear, transparent or comprehensive.

Target 12 Recommendations

The Department should work with data owners on the following:

- The Department should fully investigate the differences between NEET and Connexions guidance and advice service estimates and report on the outcome.
- The Department should gain access to student numbers at pupil referral units.
- The SFR should include a range of additional information, such as: a link to ONS estimates of sampling error; the change from ISR to ILR; the NEET calculation in further detail; a small table showing the NEET figure in total and split by gender.
- The Department should consider the level of rounding applied to the final figures, in the light of the use to which they will be put and the sampling errors.
- A December snapshot for ILR data collection should be considered.
- The APR and Departmental Annual Report should include an annex with more detail on the LFS-specific errors and the likelihood of error due to proxy responses.
- The up-to-date Memorandum of Understanding (MOU) between HESA and DfES should be expedited.
- A formal agreement over the ILR should be drafted and implemented by DfES and LSC, and made available for consultation by those using ILR data.
- Publicly-accessible documentation on data quality should be available in conjunction with publication of the NEET figure, possibly through a link to a central DfES site covering quality of all data sources used to measure PSA targets.
- Some detail about the data sources, their coverage and their roles in the NEET calculation could be included in the annual reports.
- The Connexions MI result should be disclosed with the NEET figure, together with an explanation of the differences, where known (see first bullet).
- The Technical Note should be updated with a number of corrections and expansions, including indicating that the large error margin requires the final NEET figure in 2012 to be 3ppts below the baseline in order to be sure that the target was met.

PSA Target 13

Increase the number of adults with the skills required for employability and progression to higher levels of training through:

i. Improving the basic skills levels of 2.25 million adults between the launch of Skills for Life in 2001 and 2010, with a milestone of 1.5 million in 2007;
ii. Reducing by at least 40% the number of adults in the workforce who lack NVQ2 or equivalent qualifications by 2010. Working towards this, one million adults in the workforce to achieve level 2 between 2003 and 2006.

Data Systems:

Part (i):
Individualised Learner Record (ILR)
Offender Learning and Skills Unit (OLSU) learner data
Jobcentre Plus learner data

Part (ii):
The Labour Force Survey (LFS) , published by the Office for National Statistics (ONS).

Summary of review

Part (i):

Each data source counts all learners rather than those progressing, meaning adjustments are required and the system is not relevant. Data collection methods and adjustments differ by data source, implying that the system is not well-defined. Despite potential inconsistencies between data sources, the system is largely robust. However, due to problems mentioned above, the system is not entirely reliable.

The system is largely comparable. Where adjustments have been made in the light of new analysis, adjustments have been applied retrospectively. It is also largely verifiable, with validation procedures applied to ILR and OLSU data (comprising 99% of the total).

The reporting process does not include details of how much of the data is supplied by each source and how the progress towards the target is then derived. For this and other reasons the reporting is not clear, transparent or comprehensive. The Technical Note needs to be clear on what is measured by each source, and how the final figures are calculated. Details should be given on the adjustments applied to the data from each source.

Part (ii):

The system is largely relevant, despite a risk due to LFS not sampling communal establishments. The LFS was primarily designed to provide labour force data and not to measure respondents' qualifications. Therefore a risk exists that the information obtained through the LFS is not accurate and the system is not entirely well-defined.

Problems with proxy measures, non-response bias, 'other' qualifications and reliance on respondents' recall imply that the system is not entirely robust. For the same reasons, the system may not be entirely reliable. Some back-calculations to allow for changes in the system have been carried out, and despite these changes it remains broadly comparable. DfES methodology to apportion qualification levels using a combination of the highest qualification variable and other qualification variables is not properly documented, and the system is not verifiable.

DfES report in terms of progress towards the total of 3.6 million extra level 2s required by 2010, as described in the Technical Note. The Treasury reports progress towards the interim target of 1 million extra level 2s between 2003 and 2006, and the total reduction in the number in the workforce without a level 2, and because of this inconsistency reporting is not entirely clear. The methodology behind the apportionment of 'other' qualifications is not set out in the SFR, and hence reporting is not sufficiently transparent. The reporting does appear to be comprehensive, however, as all aspects of the data system relevant to the target are included.

The Technical Note is generally satisfactory, but more disclosure of the derivation should be given.

Target 13 Recommendations

The Department should work with data owners on the following:

(i) ILR, OLSU and Jobcentre plus

- We would recommend that a new learner database to cover OLSU data be established as soon as possible.
- The annual progress towards the target should be published in an SFR.
- Details of how the target measure is derived from each of the data sources should be given when the figures are released.
- The technical note needs to be clear on what is measured by each source, and how the final figures are calculated. Details should be given on the adjustments applied to the data from each source.

(ii) LFS

- The Department should consider whether there is any other data source that could be used to measure this target.
- There should be follow-up on the ONS work to improve sample frame coverage.
- The Department should continue to monitor changes which may affect the measure, and carry out necessary back-calculations to allow for these.
- The Department should produce a document which sets out how the highest qualification is calculated using other LFS variables, and setting out clearly caveats related to proxy measures and 'other' qualifications.
- Progress reporting should be made consistently by both the Department and Treasury.
- Details of the apportionment of "other" qualifications should be disclosed in the SFR and the Technical Note, as well as information on the apportionment of

apprenticeships, Scottish Certificate of Six Years Studies (CSYS) and those responding “don’t know”.

PSA Target 14

By 2010, increase participation in higher education towards 50% of those aged 18 to 30 and also make significant progress year on year towards fair access, and bear down on rates of non-completion.

Data Systems:

HESA records, LSC records (Further Education ILR).

Summary of review

The review focuses on the calculation of the Higher Education Initial Participation Rate (HEIPR) which is used to measure the participation part of this target. The other parts of the target, fair access and non-completion are reported by DfES and use HESA performance indicators. However, these parts of the target have yet to be benchmarked and it was not possible to undertake any meaningful assessment of the data system. The summary of conclusions below is, therefore, confined to the participation part of this target and the calculation of the HEIPR.

The data system is largely relevant, but subject to undercounting due to the exclusion of certain groups of HE participants. It is also well-defined apart from the problems of over- or under-counting caused by multiple datasets. For the same reasons the system is judged to be largely robust and reliable. Effective controls operate for comparability, i.e. retrospective revisions are made to HEIPR when population counts are updated. The system is complex but internally verifiable, with the exception of the correction for false matches created when linking to prior HESA records to check for previous HE experience. Overall, it is therefore judged to be not entirely verifiable.

HEIPR is now a national statistic and is issued as a SFR in the April of each year. It includes definitions which details what is and is not included within the calculation of HEIPR. With a few caveats, the reporting is judged to be largely clear, transparent and comprehensive. The Technical Note outlines the general calculation but fails to precisely define what is a ‘nationally recognised awarding body’. No detail is given on the derivation of the numerator from multiple data sources nor any estimate of sources of error.

Target 14 Recommendations

The Department should work with data owners on the following:

- The Department should develop a comprehensive system for estimating the effects of non-response, over- and under-counting on the key statistic.
- HE in FE numbers in Wales, Northern Ireland and Scotland should be included.
- Overcounting due to failure to record prior HE in FE and double counting across LSC and HESA systems should be estimated and applied to the HEIPR.
- The correction for false matching work should be repeated.
- The fair access and non-completion parts of the target should be benchmarked.
- Estimates of overall error should be published alongside the HEIPR.
- Clearly record in the Technical Note and the SFR what is required to reach the target.
- The Technical Note should be amended to include details of calculations and references for both the numerator and denominator in detail, with sources or error and an overall estimate of error. Further links and clarifications should also be included.

Appendix 1: Review of PSA Target Data Sources – Statistical Checklist

	Question	Record in Report Template	Routing
1	Describe exactly the population relevant to the target.		
2	Is the data derived directly from the target population?	1.1: ***Relevant***	If yes, go to 4
3*	Describe the relationship between target population and population from which data collected. What is done to ensure the collected data is relevant to the target? Could changes to the target specification improve the match while maintaining the underlying policy thrust of the target?	1.2: ***Well-defined***	
4	Is the data drawn from the whole population (census) or a sample?	1.2: ***Well-defined***	If census, go to 9
5	Give details of the sampling procedure. Is the sample as drawn fully representative of the population?	2.2: ***Robust***	If yes, go to 7
6*	In what respect is the drawn sample unrepresentative? What is done to correct for this?	2.3: ***Reliable***	
7	Are statistics reported in such a way as to allow for sampling uncertainty?	2.2: ***Robust***	If yes, go to 9
8*	Give details of reporting and whether lack of reporting of sample uncertainty is likely to be critical to decision-making.	2.3: ***Reliable***	
9	Is there a likelihood of non-response bias?	2.2: ***Robust***	If no, go to 11
10*	Give details of potential non-response bias, and any ways of estimating or controlling for it.	2.3: ***Reliable***	
11	How is the outcome of interest measured, and how valid and reliable is that measurement?	1.1: ***Relevant*** 1.2: ***Well-defined*** 2.2: ***Robust*** 2.3: ***Reliable***	If validity and reliability are both sufficient, go to 13
12*	Give details of problems with validity and reliability of measurement.		

	Question	Record in Report Template	Routing
13	Over what timescale is the data collected? If it is collected periodically, what is the period (annual, every 2 years, etc.). If administrative data, how frequently is the data compiled and collected? If a moving average is used to address the target, over what period is the average computed?	1.1: ***Relevant*** 2.2: ***Robust*** 2.3: ***Reliable***	
14	Based on previous information, are the computed statistics sufficiently timely to address the target adequately?		If yes, go to 16
15*	Give details of problems with timeliness and how these might be controlled.		
16	Are there likely to have been changes over time in the measures collected or the relationships between them and the underlying quantities of interest?	2.4: ***Comparable***	If no, go to 18
17*	Give details of possible changes over time which might affect the way in which the collected data relates to the underlying quantities.		
18	What quality assurance procedures are in place to ensure that information is captured, transferred and maintained in a secure fashion, and that statistics are computed appropriately? Are these adequate?	2.1: ***Verifiable*** 2.2: ***Robust*** 2.3: ***Reliable***	If yes, go to 20
19*	Give details of potential problems with data capture, transference or maintenance and possible controls.		
20	Give details of documentation of derivation of target outcomes and data sources (Technical Note), how outcomes are reported and what details are given of their derivation. Are these all clear, transparent and comprehensive?	3.1: ***Clear*** 3.2: ***Transparent*** 3.3: ***Comprehensive***	If yes, finish
21*	Give details of problems with documentation, reporting or disclosure, and possible controls or improvements which could be put in place.		

(* - potential risk. Include details of reporting and disclosure of risks that currently operate or should be put in place)

Link to ONS framework for statistical quality, which may also be worth looking at:

http://www.statistics.gov.uk/StatBase/Product.asp?vlnk=13578

Appendix 2: Issues on Estimates Based on Sample Data

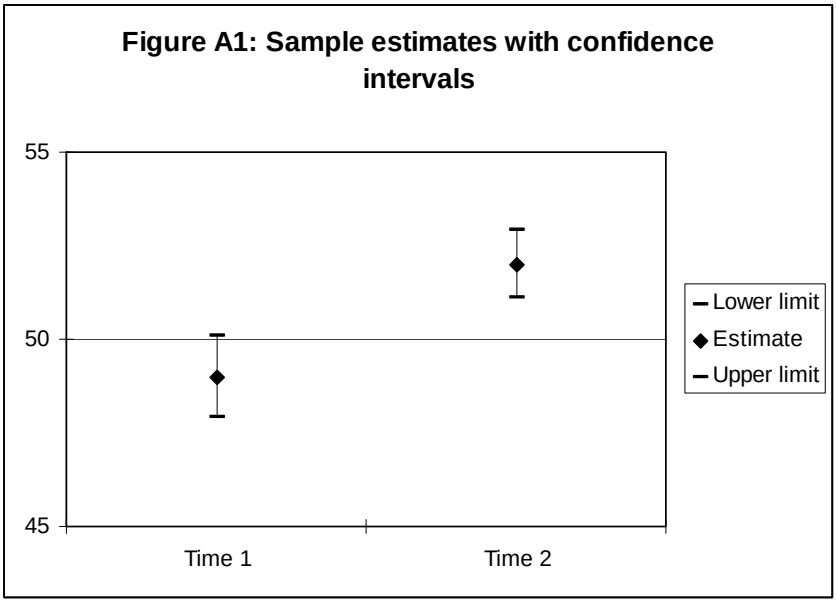
On occasions it is not practical, or cost-effective, to use a census of all individuals in a population in order to derive a measure of interest. In these cases we would use a sample, drawn from the population in a random fashion, which is intended to be representative of the underlying population and allows us to estimate the required measure to an acceptable level of accuracy. The issues which arise when samples are used include:

- Estimates from samples are always subject to a certain lack of precision, as well as potential bias in the estimate
- Comparisons between two sample estimates have to take into account the lack of precision in both.

We shall return to the important issue of bias later. For the moment, let us assume we have two sample estimates which we wish to compare – for example, estimates of an underlying population value at two points in time, from which we need to decide if that population value is really showing an upward trend. The following diagrams illustrate some of the possible scenarios.

Figure A1 shows sample estimates at two time points, each with their 95% confidence intervals. These are ranges of values within which there is a 95% chance that the true underlying population value lies. In Figure 1, the two ranges do not overlap, so we can be reasonably confident that there has been a real change in the underlying value[6].

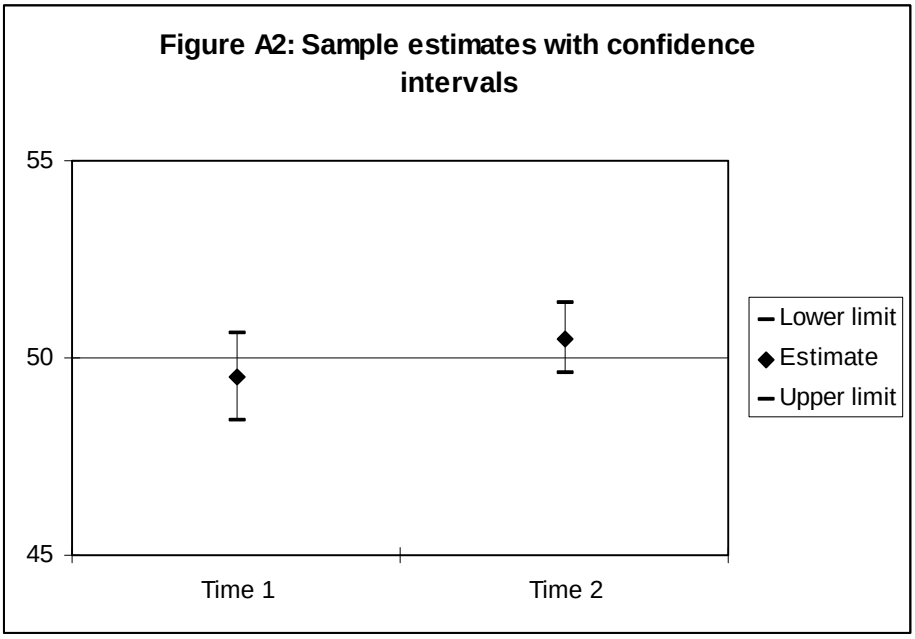
Figure A1: Sample estimates with confidence intervals



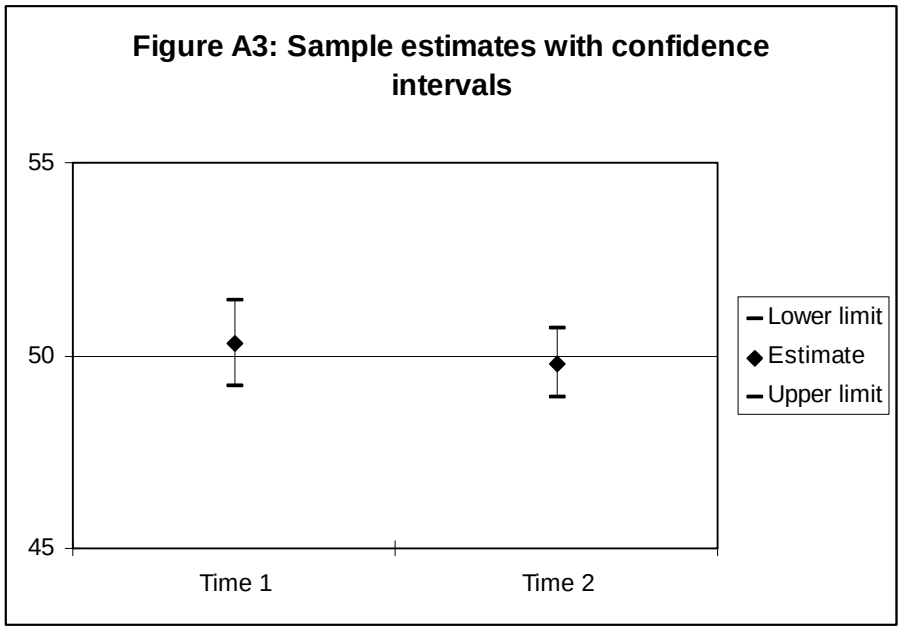
[6] Note that it is only approximately true to say that if the two ranges do not overlap there has been a real change – but it can be used as a fairly reasonable rule of thumb.

The above figure also allows us to illustrate issues to do with meeting a fixed target, rather than demonstrating change over time. If the target here was a value of 50, at Time 1 the estimate was below, but the upper limit of the confidence interval was above. From this we could say that that there was a non-zero (but fairly small) chance that the underlying population value was actually above 50. To be pretty certain it is above 50 we would need a situation like at Time 2 in Figure A1, where the whole confidence interval is well clear of the 50 line.

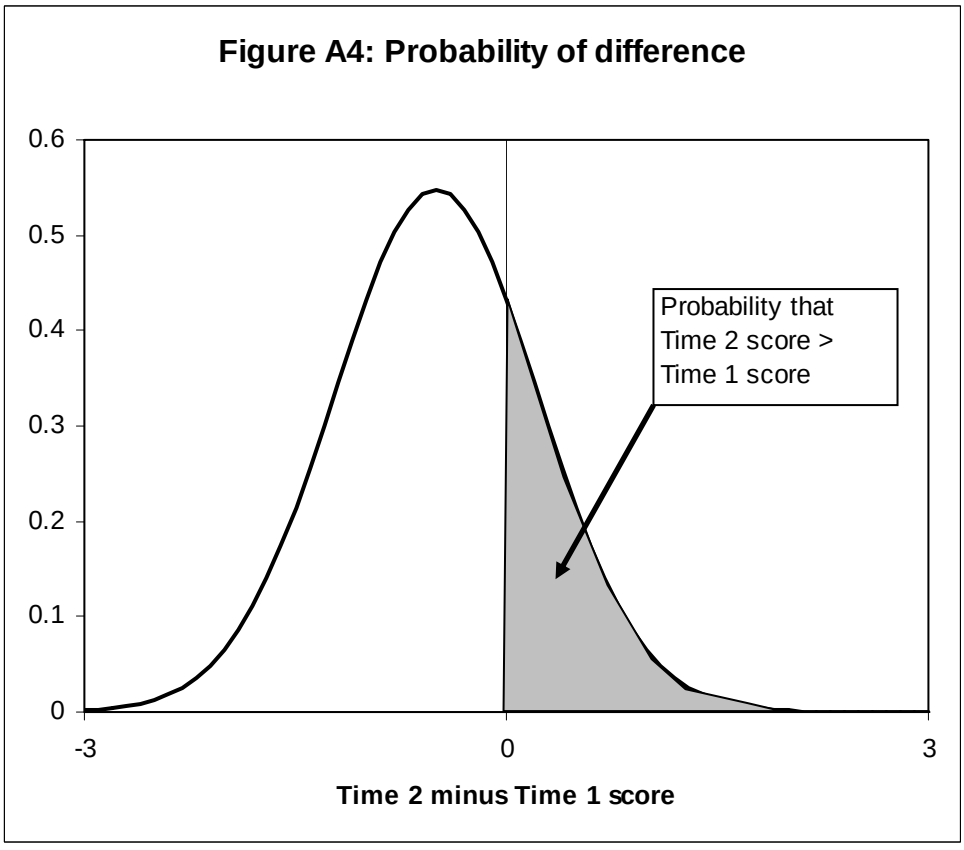
Figure A2 shows a very similar scenario, but this time the two confidence intervals overlap, implying that we cannot be confident that there has been a real rise in the underlying value even if the actual estimates themselves have risen.



Note that in neither case can we assert with confidence that the target of 50 has been met. A very similar scenario is shown in Figure A3 below, but in this case the actual estimates (indicated by diamonds) show a decline.



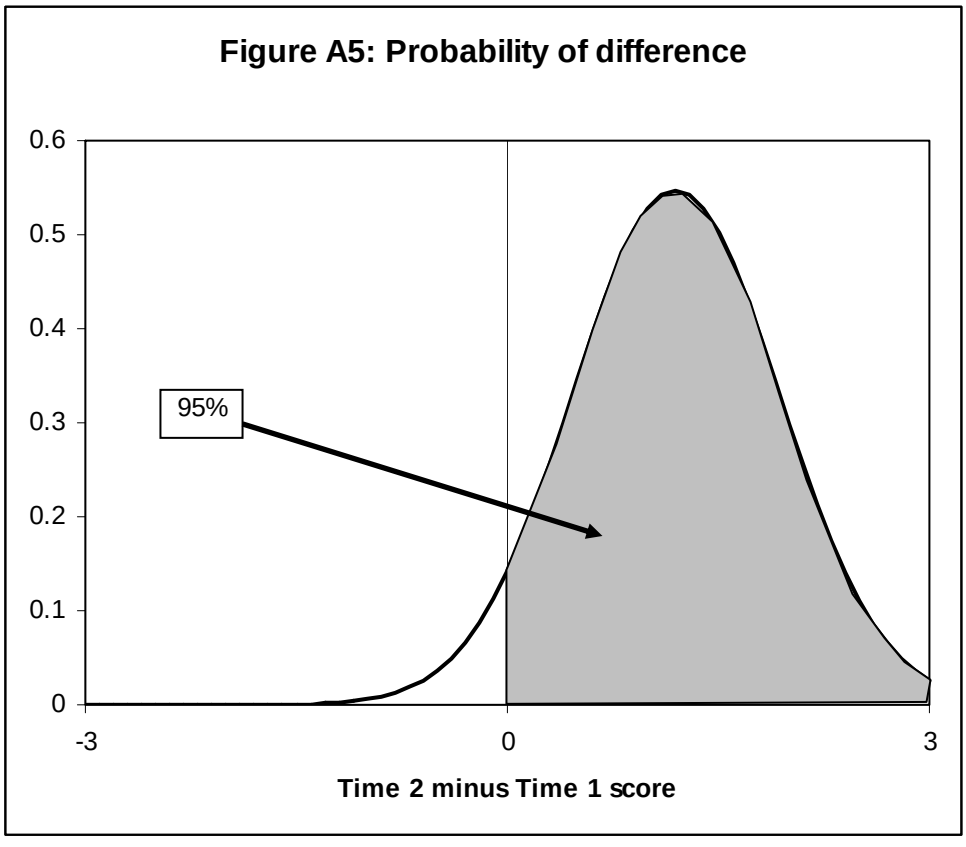
We can go on to investigate the probability that there has actually been a rise from Time 1 to Time 2 in the underlying population value, using a probability distribution figure like Figure A4 below, for the different possible values of the difference between Time 2 and Time 1.



The difference between the underlying scores at Time 2 and Time 1 can take on a range of values, with probabilities as shown above based on the sample estimates we have obtained. The shaded area shows the total probability that the Time 2 value is actually greater than the Time 1 value – this probability is 25% in the above example.

How big do we want this probability to be before we can say with confidence that there has been a real improvement from Time 1 to Time 2? To some extent, this is a matter for policy makers to decide, but a statistician might advise that 95% was a good figure, giving only one chance in 20 of incorrectly deciding there had really been a positive change when this was not true. For this level of confidence, we would need a plot looking something like Figure A5 below.

The calculation of these sampling uncertainties and working out the probability that there has been a real change, or that the target has really been met, are non-trivial exercises but ones which can be carried out from basic sampling information by statisticians. This should be part of the routine when performance against targets is being evaluated, and should be taken into account when decisions are being made about sample sizes and other related aspects of data collection.



In addition to the concerns about precision of estimates which have been outlined above, the other issue which should be considered is bias. Bias occurs when there are systematic differences between the population characteristics and those of the sample, for example when significant groups are excluded from the sampling. The characteristics of such excluded groups should be considered, and an objective assessment made of any likely impact on the sample estimates.

In the remainder of this appendix we give more detailed information about sampling, for the benefit of those who wish to understand this area, and the concepts of bias and precision, in more detail.

Further details on sampling

There are two aspects of samples which relate to how well they estimate the underlying characteristics of the population from which they are drawn. Both can be thought of as 'average' properties of samples, assuming we keep going back and drawing different samples from the same population, and then compared their results with each other and with the population.

Bias is a measure of the extent to which the samples accurately estimate the population properties, on average. ***Precision*** is a measure of the extent to which the results of different samples agree with each other. We shall consider both these characteristics in more detail below, and show how they relate in particular to the use of sample data to measure progress towards PSA targets.

1. Bias

Ideally, bias should be zero. Some samples may over-estimate and others under-estimate, but on average they should get it right. If the bias is very different from zero, then we have a serious problem with our sampling. The existence of bias may be related to our choice of ***sampling frame.***

A sampling frame is a list or database which contains all the members of the desired population from which the sample is to be drawn. Clearly it is an essential prerequisite for successful sampling – if the sampling frame is incomplete or flawed, then the sample may be unrepresentative of the population. The wrong choice of sampling frame can lead to big errors. There are examples in history when using a telephone survey (sampling frame = phone book) to estimate voting intentions led to a great deal of bias, as poorer voters without phones were excluded. The quality of the sampling frame should be assessed in any evaluation of a sample-based data system.

Non-response bias refers to differential response rates to a survey from different groups of individuals, or more generally to missing data being unrepresentative of the

population (e.g. traveller children not fully represented in a sample of school-age pupils). It is possible to estimate the effect of this and correct for it using weighting, provided sufficient information is available about relevant sub-groups of the population and their response rates.

2. Precision

One of the determining factors here is the size of the sample; very roughly, the precision is related to the square root of the sample size. If we want to double the precision, then we need a sample approximately four times as big. However, this is also affected by the ***design effect***, which is a number that expresses the 'efficiency' of the sample relative to the simple case where the sample is drawn purely randomly from the sampling frame with no clustering or stratification.

Given a sample size and a value of the design effect[7], it is simple to compute the precision of the sample estimate and convert this to a confidence interval. For example, if a sample of 10,000 is drawn, with a design effect of 1.0, then the precision with which a population percentage of around 50% is estimated can be shown to be plus or minus 1 percentage point, with a 95% confidence interval. In other words, if our estimate is 48.6% then we are 95% sure that the true population value is between 47.6% and 49.6%. This can clearly affect the chances of knowing whether a particular target has been met or not.

The calculation or estimation of such confidence intervals is a crucial part of using sample estimates, and should be routine for any data system which does not use a complete census of the population. If the computed confidence interval is too wide, then it can be reduced by increasing the sample size. For example, an interval which is plus or minus half a percentage point could be achieved in the above example by specifying a sample size of 40,000 cases.

Tables 1 and 2 at the end of this appendix give indications of required sample sizes for different degrees of precision, assuming simple random sampling (design effect = 1). Table 1 deals with the case where we are estimating a proportion of the order of 50%, and the required precision of the estimate is given as percentage points either side of the sample estimate for the 95% confidence interval. Table 2 deals with the case where the estimate is of a numerical value (e.g. average score) and the precision is expressed as a percentage of the underlying population standard deviation.

In summary, we can deal with low precision by increasing the sample size. If a sample is biased, however, this is a fundamental problem which needs to be dealt with by a review of the way in which the sample is being drawn.

From the above brief discussion, it is clear that there are a number of considerations to be taken into account when designing a sample system for estimating population measures. However, sampling theory allows us to address

[7] Design effects can be estimated from the properties of the sample; this is not always straightforward, but statisticians can usually make reasonable estimates from previous data.

these issues and there is no reason why a well-designed sample should not give unbiased results with a degree of precision which can be calculated and quoted.

Using sample estimates to compute the probability of having reached a Target

Suppose we have estimated a population parameter X with a sample from the population, giving us an estimate E with a standard error S – in other words, the 95% confidence interval for the true value of X is between E-1.96S and E+1.96S. If we have a target value for X, T say, what is the probability that the target has been reached?

P[X >= T] = P[Z >= (T-E)/S],

Where Z is a Standard Normal variable.

This is 1 - Φ((T-E)/S), where Φ() is the Standard Normal Integral Function.

For example, if the Target value is 50 and the sample estimate is 49 with a standard error of 2 units, then the probability that the target has actually been passed is:

1 - Φ((50-49)/2) = 1 - Φ(0.5) = 0.31 or 31%.

Alternatively, if the estimate were 52 with the same standard error, then the probability that the target had been passed would be:

1 - Φ(-1) = 0.84 or 84%.

In these situations, it might be appropriate to determine a minimum probability (e.g. 90% or 95%) in order to be able to say with confidence that a Target has been passed.

When the Target relates to a *change* from one time to the next, then the same principles apply, except that if both estimates are sample-based then the combined standard error of the difference needs to be taken into account.

Further reading

BARNETT, V. (1981) **Elements of Sampling Theory.** Sevenoaks: Hodder & Stoughton.

HAMMOND, P. (2006) 'Tool-kit 2: an introduction to sampling', in *Practical Research for Education, Issue 35, May 2006*, pp. 5-12.

KISH, L. (1965) **Survey Sampling.** New York: Wiley.

ONS: **Sample Design and Estimation.**
http://www.statistics.gov.uk/about/data/methodology/general_methodology/sde.asp

Table 1: Simple Random Sample Sizes for Precision of Estimating Proportions

Required Precision (percentage points)	Simple Random Sample Size
1	10,000
2	2500
3	1111
4	625
5	400
7.5	178
10	100
20	25

Table 2: Simple Random Sample Sizes for Precision of Estimating Averages

Required Precision (% of standard deviation)	Simple Random Sample Size
1%	40,000
2%	10,000
3%	4444
4%	2500
5%	1600
7.5%	711
10%	400
20%	100

Appendix 3:
References and Documentation for each SR2004 Target Data System Review

PSA Target 1	
1.1	http://www.standards.dfes.gov.uk/primary/publications/foundation_stage/940451/qca_fs_profilehbk_031006.pdf
1.2	http://www.qca.org.uk/downloads/5585_cg_foundation_stage.pdf
1.3	'Further Analysis of Index of Multiple Deprivation 2004 for targeting and monitoring Sure Start policies/resources' by Stefan Noble, Gemma Wright & Michael Noble (University of Oxford)
1.4	http://www.dfes.gov.uk/rsgateway/DB/SFR/s000476/SFR25-2004.pdf
1.5	http://www.dfes.gov.uk/rsgateway/DB/SBU/b000537/sfr44-2004v2.pdf
1.6	'Implementation and moderation of foundation stage profile assessment 2005 annual monitoring report' by the NAA (still in draft form)
1.7	'Customer requirements specification: Foundation stage profile 2006-aggregated data'
1.8	'2006 Foundation stage profile (FSP) software specification' document
1.9	PUPIL03 spreadsheet
1.10	http://www.dfes.gov.uk/rsgateway/DB/SFR/s000458/10-2004v2.pdf

PSA Target 2	
2.1	http://www.ofsted.gov.uk/daycare/
2.2	http://www.dfes.gov.uk/research/data/uploadfiles/RR348.pdf
2.3	http://www.dwp.gov.uk/
2.4	Calculation of null provider estimate provided by OFSTED
2.5	http://www.dwp.gov.uk/asd/frs/2003_04/methodology/non_response.asp
2.6	PSA Validation Summary for DWP SR2002 PSA (HBAI)
2.7	OFSTED desk instructions, etc.
2.8	http://www.ifs.org.uk/wps/wp0123.pdf
2.9	DfES Autumn Performance Report 2005, Cm 6719
2.10	Comments on the review from stakeholders

PSA Target 3	
3.1	Conception Statistics for 2002 with notes
3.2	Birth Statistics 2003 with notes
3.3	2004 Abortion Statistics and notes
3.4	Teenage Pregnancy Unit - Statistics and notes
3.5	Health Statistics Quarterly, No.10, Summer 2001
3.6	Information on population estimates
3.7	Population estimates methodology and sources
3.8	Documentation on the imputation techniques used in the processing of births data. (Confidential - access restricted). Contact Nic Gillam (ONS).
3.9	The SLA between ONS and DH detailing the imputation techniques to be used in the processing of abortion data. (Confidential - access restricted). Contact Nirupa Dattani (ONS)
3.10	Statistical Bulletin – Abortion Statistics, England and Wales: 2004, Department of Health 2005
3.11	Birth Statistics, Historical Series, 1837-1983, Series FM1 no13
3.12	Documentation regarding the checking procedures in fertility statistics. (Confidential - access restricted). Contact Nirupa Dattani (ONS)
3.13	National Statistics methodology
3.14	About National Statistics/Code of Practise
3.15	Documentation on the suppression of ward conception data. (Confidential - access restricted). Contact Nirupa Dattani (ONS)

PSA Target 4	
4.1	http://www.dh.gov.uk/assetRoot/04/09/89/12/04098912.pdf §2.
4.2	www.statistics.gov.uk/statbase/ssdataset.asp?vlnk=9085&more=y
4.3	www.dh.gov.uk/publicationsandstatistics/publishedsurvey/healthsurveyforengland/fs/en
4.4	http://www.dh.gov.uk/assetRoot/04/09/89/12/04098912.pdf pp. 159-163.
4.5	http://www.dh.gov.uk/assetRoot/04/09/89/12/04098912.pdf §2.1
4.6	http://www.dh.gov.uk/assetRoot/04/09/89/12/04098912.pdf §2.2
4.7	www.data-archive.ac.uk/findingdata/sndescription.asp?sn=5098
4.8	http://www.archive2.official-documents.co.uk/document/deps/DH/survey02/md/md-01.htm §7.2.
4.9	As [4.5], §8.7 and tables 12 to 25.
4.10	As [4.5], §7.2, §7.3, §7.4.
4.11	http://www.dh.gov.uk/assetRoot/04/10/94/10/04109410.pdf
4.12	http://www.archive2.official-documents.co.uk/document/deps/DH/survey02/hcyp/hcyp31.htm §9.1
4.13	http://www.dfes.gov.uk/SR2004/PSA/SR04_Technical%20Notes_Online_29-Jul-04.doc
4.14	NatCen Code of Practice and quality assurance documentation. (Confidential - access restricted). Contact

PSA Target 5	
5.1	http://www.dfes.gov.uk/datastats1/guidelines/children/returns.shtml See OC2 and SSDA903.
5.2	DfES Autumn Performance Report 2005, Cm 6719

PSA Target 6	
6.1	http://www.qca.org.uk/eara/337.html
6.2	http:www.qca.org.uk/12331.html
6.3	http://www.qca.org.uk/eara/452.html
6.4	http://www.qca.org.uk/eara/314.html
6.5	http://www.dfes.gov.uk/rsgateway/DB/SFR/s000564/index.shtml
6.6	Copies of internal analysis documents
6.7	http://www.dfes.gov.uk/rsgateway/DB/SFR/s000564/index.shtml
6.8	http://www.qca.org.uk/downloads/6301_compare_nat_tests_96-01.pdf
6.9	http://www.statscom.org.uk/media_pdfs/reports/023_-_Measuring_Stds_in English schools.pdf
6.10	http://www.dfes.gov.uk/performancetables/primary_05/p4.shtml
6.11	http://www.dfes.gov.uk/performancetables/primary_05/glossary.shtml
6.12	Comments on the review from stakeholders

PSA Target 7	
7.1	http://www.qca.org.uk/eara/337.html
7.2	http:www.qca.org.uk/12331.html
7.3	http://www.qca.org.uk/eara/452.html
7.4	http://www.qca.org.uk/eara/314.html
7.5	http://www.dfes.gov.uk/rsgateway/DB/SFR/s000564/index.shtml
7.6	Copies of internal analysis documents
7.7	http://www.dfes.gov.uk/rsgateway/DB/SFR/s000564/index.shtml
7.8	http://www.qca.org.uk/downloads/6301_compare_nat_tests_96-01.pdf
7.9	http://www.statscom.org.uk/media_pdfs/reports/023_-_Measuring_Stds_in English schools.pdf
7.10	(Need refs for production of secondary AAT)
7.11	As above
7.12	Comments on the review from stakeholders
7.13	http://www.qca.org.uk/downloads/QCA-05-1874-KS3ICT05pilot.pdf
7.14	http://www.dfes.gov.uk/rsgateway/DB/SFR/s000599/index.shtml

PSA Target 8	
8.1	..\..\Data System Docs\Stakeholder Comments\Specification comments 1 KSear.htm ..\..\Data System Docs\Stakeholder Comments\Speification comments 2 KSear.htm ..\..\Data System Docs\Stakeholder Comments\Target comments SQuazi.doc
8.2	..\..\Data System Docs\REGISTRATION CODE GUIDANCE v1.51.doc
8.3	..\..\Performance Docs\E-REG 2ND INTERIM REPORT JAN05.doc
8.4	L:\PSATargets\Target 8\documentation\HC 212.pdf
8.5	..\..\Data System Docs\REGISTRATION CODE GUIDANCE v1.51.doc
8.6	..\..\Data System Docs\Stakeholder Comments\target comments 2 SQuazi.doc ..\..\Data System Docs\Stakeholder Comments\Target comments SQuazi.doc
8.7	..\..\Data System Docs\validationdoc.doc ..\..\Data System Docs\Sample Validation report from FORVUS.doc
8.8	..\..\Data System Docs\production rules for SFR.doc
8.9	..\..\Data System Docs\Stakeholder Comments\Frame cleaning comments PRobinson.htm ..\..\Data System Docs\Stakeholder Comments\Validation comments KSear.htm
8.10	..\..\Data System Docs\SCDB Data Dictionary Extract.doc
8.11	..\ONS Data Quality Assessment Guidelines.pdf
8.12	L:\PSATargets\Target 8\documentation\School Census Preparation and Guidance 2006 Version 1.1doc.doc
8.13	..\..\Data System Docs\REGISTRATION CODE GUIDANCE v1.51.doc
8.14	..\..\Data System Docs\response rate analysis.xls
8.15	..\..\Data System Docs\non-repsonse bias\all reg.spo
8.16	..\..\Data System Docs\validationdoc.doc
8.17	..\..\Data System Docs\production rules for SFR.doc
8.18	http://www.statistics.gov.uk/about/data/methodology/quality/
8.19	http://www.hm-treasury.gov.uk/performance/dfes.cfm
8.20	..\..\Performance Docs\SFR56-2005.pdf
8.21	http://www.dfes.gov.uk/SR2004/PSA/SR04_Technical%20Notes_Online_6-Aug-04.doc
8.22	..\..\Performance Docs\2005deptrep.pdf
8.23	..\..\Performance Docs\SFR56-2005.pdf
8.24	..\..\Data System Docs\forvus absence form.pdf

(NB: Many of the above links are non-operational and need to be regenerated or the documents made accessible in a suitable form)

PSA Target 9	
9.1	http://www.teachernet.gov.uk/teachingandlearning/subjects/pe/penews/200405_School_Sport_Survey/
9.2	2002/03: http://www.archive2.official-documents.co.uk/document/deps/ofsted/170/ 2003/04: http://www.ofsted.gov.uk/publications/annualreport0304/ 2004/05: http://www.ofsted.gov.uk/publications/annualreport0405/
9.3	http://www.ofsted.gov.uk/publications/annualreport0304/subject_reports/secondary/pe.htm http://www.ofsted.gov.uk/publications/annualreport0304/subject_reports/primary/primarype.htm
9.4	The School Sport Partnership Evaluation 2005: Key quality procedures
9.5	http://www.dfes.gov.uk/SR2004/PSA/SR04_Technical%20Notes_Online_6-Aug-04.doc
9.6	http://www.dfes.gov.uk/publications/deptreport2005/

PSA Target 10	
10.1	'2005 Matching Contractor – Base and Change Control Specification'
10.2	Data Requirement sent to awarding bodies by contractor
10.3	Section 96
10.4	QCA Code of Practice
10.5	QCA reports into specific GCSE subjects
10.6	A review of GCE and GCSE coursework arrangements
10.7	The Department PSA Technical Note
10.8	SFR 25/2005
10.9	The Department Autumn Performance Report and Annual Report
10.10	DfES Schools and Colleges Achievement and Attainment Tables
10.11	Comments on the review from stakeholders

PSA Target 11	
11.1	http://www.statistics.gov.uk/nsbase/methods_quality/quality_review/downloads/Report_education38.doc
11.2	http://www.dfes.gov.uk/rsgateway/DB/SFR/s000561/index.shtml
11.3	http://www.statistics.gov.uk/methods_quality/quality_review/downloads/MeasurePlan.doc
11.4	http://www.dfes.gov.uk/SR2004/PSA/SR04_Technical%20Notes_Online_6-Aug-04.doc
11.5	Department Annual Report
11.6	Department Autumn Performance Report

PSA Target 12	
12.1	http://www.dfes.gov.uk/rsgateway/DB/SFR/s000587/index.shtml
12.2	http://www.dfes.gov.uk/publications/deptreport2005/docs/autumn2k5annualreport.pdf
12.3	http://www.dfes.gov.uk/publications/deptreport2005/docs/2005deptrep.pdf
12.4	http://www.dfes.gov.uk/SR2004/PSA/SR04_Technical%20Notes_Online_6-Aug-04.doc
12.5	PLASC questionnaire - (Contact: Jim Foley, James.Foley@dfes.gsi.gov.uk, ext. 62648)
12.6	Connexions Management Information - (Contact: Gill Cowan, Gillian.Cowan@dfes.gsi.gov.uk, ext, 23117)
12.7	Comparison of population estimates with schools census information - (Contact: Phil Rose, Phil.Rose@dfes.gsi.gov.uk, ext. 24086)
12.8	http://www.statistics.gov.uk/downloads/theme_labour/PQMD04F.pdf
12.9	Report on standard error, standard deviations and data definitions - (Contact: Phil Rose, Phil.Rose@dfes.gsi.gov.uk, ext. 24086)
12.10	http://www.statistics.gov.uk/STATBASE/Source.asp?vlnk=358&More=Y
12.11	http://www.statistics.gov.uk/downloads/theme_population/Making_PopulationEstimate.pdf
12.12	http://www.statistics.gov.uk/census2001/annexa.asp
12.13	PLASC quality document - (Contact: Mike Battle, Mike.Battle@dfes.gsi.gov.uk, ext. 62038)
12.14	HESA Memorandum of Understanding - (Contact: Stephen Cook, Stephen.Cook@dfes.gsi.gov.uk, ext. 62343)
12.15	ILR data access protocols - (Contact: Michael Greer, Michael.Greer@dfes.gsi.gov.uk, ext. 24029
12.16	Youth 4 Participation documentation - (Contact: Phil Rose, Phil.Rose@dfes.gsi.gov.uk, ext. 24086)
12.17	Product description: Schools data - (Contact: Phil Rose, Phil.Rose@dfes.gsi.gov.uk, ext. 24086)
12.18	Product description: FE data - (Contact: Phil Rose, Phil.Rose@dfes.gsi.gov.uk, ext. 24086)
12.19	Product description: HE data - (Contact: Phil Rose, Phil.Rose@dfes.gsi.gov.uk, ext. 24086)
12.20	Comparison of provisional, revised and final data - (Contact: Phil Rose, Phil.Rose@dfes.gsi.gov.uk, ext. 24086)
12.21	http://www.dfes.gov.uk/rsgateway/DB/SFR/s000551/index.shtml
12.22	http://www.dfes.gov.uk/rsgateway/DB/SFR/s000568/index.shtml
12.23	http://www.dfes.gov.uk/rsgateway/DB/SFR/s000637/index.shtml
12.24	Comments on the review from stakeholders

PSA Target 13.1	
13.1.1	ILR questionnaire
13.1.2	OLSU questionnaire
13.1.3	http://pso.hmprisonservice.gov.uk/PSO_7100_kpi_and_kpt_guidance_2004-5.doc
13.1.4	SPSS syntax to extract FE F04 ILR data for the September 2005 local basic skills statistical update
13.1.5	ILR credibility report
13.1.6	http://www.hm-treasury.gov.uk./performance/targets/perf_target_8.cfm
13.1.7	Comments on the review from stakeholders

PSA Target 13.2	
13.2.1	DfES PSA Target 13 Questionnaire
13.2.2	http://www.statistics.gov.uk/StatBase/Product.asp?vlnk=1537&Pos=&ColRank=1&Rank=272
13.2.3	MORI Improving Quals ITT
13.2.4	SPR04 notes to LFS users
13.2.5	Hiqual level template (to 2004)
13.2.6	http://www.hm-treasury.gov.uk./performance/targets/perf_target_9.cfm
13.2.7	http://www.dfes.gov.uk/rsgateway/DB/SFR/s000562/index.shtml
13.2.8	Comments on the review from stakeholders

PSA Target 14	
14.1	http://www.dfes.gov.uk/rsgateway/DB/SFR/s000572/index.shtml
14.2	http:\\www.statistics.gov.uk/methods_quality/quality/quality_review/education.asp
14.3	Email from Mark Gittoes (HEFCE) to Kathryn Kelly (DfES) 9 Dec 2005
14.4	http://www.hesa.ac.uk
14.5	Department presentation pack (informal Powerpoint presentation not intended for publication)
14.6	Comments on the review from stakeholders

Appendix 4: Notes on Review of SR2002 Targets

The detailed review was carried out by a team of DfES statisticians:

Jon Andrews
Julian Austin*
Andy Brook
Natalie Corke
Michael Greer
Sachin Patel
Kate Shaw
Ben Stanbury
Ian Thomson
Gill Turner
Ann Wass,

in collaboration with NAO examiners:

Mo Choudhury
Sarah Edwards
Anthony Goreham
Jo Harris
Sascha Kiess
Anne Taylor *
Lola Toppin

In this appendix we will give a brief summary of the work done under this heading, which then fed into the full-scale review of SR2004 target data systems presented in the body of this report. DfES statisticians worked alongside NAO staff to respond to issues raised by NAO on the SR2002 targets, and to ensure that as complete a picture as possible of the underpinning data systems was presented.

This summary begins with a description of the procedure followed to reach the final judgement about each target's data system, including quality assurance measures which were inherent in the process. A description of the NAO's validation system is also included. An overall description of the situation regarding the SR2002 data systems at the end of this review process is also given.

Procedures and Quality Assurance Measures

The following procedures were followed in order to derive the sections of this report, focusing on the data systems underpinning each of the SR2002 PSA targets:

1. NAO produced interim assessments for each target and raised a number of issues.

* Julian Austin and Anne Taylor only worked on 2004 targets

2. A DfES statistician was allocated to each target, alongside a member of NAO personnel, and given the responsibility to ensure that the interim conclusions about the target were updated in line with the best available evidence.
3. An external statistical consultant was appointed to advise on and coordinate the process.
4. A list of questions was drawn up for each target to cover all outstanding issues for which an update was required.
5. The external consultant met with each of the statisticians to discuss issues relating to their target.
6. For each target the statistician drew up responses to the questions, based on the best available evidence they were able to obtain, and shared these with the NAO person to ensure agreement.
7. These responses were sent to the external consultant, who drew up a draft report for all targets in a consistent format and sent it to the statisticians for comment.
8. Comments from statisticians were incorporated in a second draft report which was considered by the project Steering Committee on 14th December 2005.
9. This final version was produced in the light of modifications and enhancements requested by the Steering Committee.

The above process incorporated the following quality assurance elements:

- A focus on responses to specific issues raised by NAO.
- Healthy debate between NAO person and DfES statistician on each data system.
- Meetings between the external consultant and each statistician to ensure that quality issues had been explored thoroughly.
- Final agreement in each case between DfES and NAO on the appropriate validation ratings for each target.

NAO's Validation System

The validation ratings awarded to data systems underpinning a department's PSA targets are described as follows:

Validation rating	**Description**
White	The system is not sufficiently established to form a view on its fitness for purpose
Green	The data system is fit for purpose and effective controls have operated
Amber	The data system addresses the majority of risks to data quality but: (i) needs strengthening to ensure that remaining risks are adequately controlled; or (ii) includes limitations that cannot be cost-effectively controlled; the Department needs to explain the implications of these more clearly to the reader
Red	The data system is not fit for the purpose of measuring and reporting performance against the target
Not rated	The target has since been superseded and no progress against it could be measured

NAO's flowchart which captures the decision processes underpinning the award of the above ratings is given in the main report.

Summary and recommendations

The table below summarises the SR2002 target data systems assessed in this report, the NAO draft validation ratings awarded in their initial assessment, and any changes that might be recommended on the basis of the work reported here:

SR2002	NAO Interim validation rating on SR2002 target	NAO Updated validation rating (all agreed by DfES statisticians)	SR2004 Target (* - closely equivalent; ** - approximately equivalent)
1	Green	Green	6*
2	Green	Amber	7*
3	Amber	Amber	8**
4	Red	Amber	9*
5	(i) Amber, (ii) Red	(i) Amber, (ii) Red	10*/11*
9	(i) Amber, (ii) & (iii) White	(i) Amber, (ii) & (iii) Red	14
10	Amber	Amber	13(i)*
11	Amber	Amber	13(ii)
SS 1	White	Not rated	1*
SS 2	White	Red	
SS 3	White and Green	Not rated and Green	1*
SS 4	Amber	Amber	2**

(Note that targets 6 to 8 are not included in the current review)

The following general points have emerged in the course of this work which are relevant to the review of the SR2002 targets and also to the work on SR2004 targets and data systems.

- There are some generic issues which relate not so much to data systems but to the validity or consistency of the measures which they capture. The two most prominent of these are the question of the extent to which national tests (including GCSE exams) capture consistent standards over time, and the issue of the consistency with which quite different qualifications are decreed to count towards a particular level of attainment. Both these issues properly come within the remit of QCA, and were explored in more detail during the review of SR2004.

- Similarly, certain data systems (PLASC and ILR in particular) underpin several targets, and generic evaluations which can be multiply referenced have been produced.

- In some cases there are problems with the wording of the target itself, rather than the data system, and it should be possible to recommend changes which would ensure more easily measurable targets in future. It would seem sensible to ensure that statisticians familiar with available data systems be involved in the development of future targets.

- The use and interpretation of estimates based on sample surveys needs to be considered as a general issue, and guidelines on how valid confidence intervals should be developed and used, in the context of deciding if targets have been met, need to be developed.

- Documentation is the key to successful control of risks and effortless evaluation of data systems – if the evidence is readily available for auditors to convince them of the robustness and reliability of the system, then a lot of work and potential conflict is avoided. Technical notes should probably play a key role in this, but it seems that in many cases these are not sufficiently well developed.

Progress since NAO's interim assessment

After NAO's interim assessment, a certain amount of progress was made relative to a range of data systems underpinning various of the SR2002 targets. This is summarised in the table below. Further progress was shown in the review of the data systems for the SR2004 targets.

Target	Progress since interim SR2002 assessment
1	QCA has responded to the Statistics Commission on standards over time. Progress towards school level targets was included in the 2004 Annual Performance Report. Excluded schools have been shown not to provide a significant risk to the headline figure.
2	As above.
3	Data collection is moving to PLASC, collected termly.
4	Data collection being extended to all schools by 2006/7. Guidance on nature of 'high quality sport' being provided.
5	Exclusions explained in TN and SFR. Move to administrative data will significantly improve measurement.
9	IER replaced by HEIPR, which will provide significant improvements in measurement.
10	Little change
11	MORI survey commissioned to judge accuracy of responses to qualifications on the LFS.
SS 1	Superseded by SR2004 target 1
SS 2	Little change – not carried forward into SR2004
SS 3	Part (i) carried forward into SR2004 – 'normal levels' now defined by FSP. Part (ii) not carried forward.
SS 4	Data for 4-year-olds now provided, and APR notes that data refers to Sure Start areas only.

In summary, therefore, nine of the above targets have recorded some progress in terms of resolving issues which were raised earlier. In some cases the progress

involves improved reporting of risks, but in others it involves a wholesale move of data system from one which was not fit for purpose to one which is expected to produce a much more robust measure to address the target.

In terms of NAO's initial assessments, if we exclude those which were assessed as white, most have not changed since then. One has gone from green to amber (target 2) and one from red to amber (target 4). The former change was not because of any substantive change to the data system (in fact, things have improved slightly, as shown above) but to ensure consistency with another target. The positive change, however, was due to improvements in the data system.

There are therefore signs of progress since the initial SR2002 assessment which will need to be continued for the SR2004 and future targets to ensure that they are all fit for the purpose of giving credible assessments of progress towards those targets.