

**Review of Data Systems Underpinning DfES  
SR2004 PSA Targets**

**A further analysis of ‘amber’ data systems**

**by**

**Ian Schagen**

**November 2006**



## **‘Amber’ Data Systems – Further Analysis**

### **Introduction**

The previous report<sup>1</sup> was based on work carried out by NAO staff and Department statisticians, with the aim of deriving a classification of the various data systems underpinning PSA targets by means of a ‘traffic light’ system. Categories included ‘green’ (largely fit for purpose), ‘amber’ (needs improvement) and ‘red’ (not fit for purpose). In the event the majority (12) of the target or sub-target systems were classified as ‘amber’, with very few ‘green’ or ‘red’ (see Table 2 of the previous report).

There were two concerns which sprang from this finding. One was that the broad category of ‘amber’ did not discriminate finely enough and it was not possible to tell how many of the systems were relatively close to ‘green’ status and how many were far from fit for purpose. Related to this was a concern that the Department needed guidance as to the ease with which as many data systems as possible could be moved into the ‘green’ category, so that significant progress could be shown in response to the original review.

This brief paper is an attempt to address these two concerns. Firstly, we decompose the NAO ‘amber’ category into two sub-categories or ‘tracks’ depending on the exact wording in the report of the reason why the data system needs improvement. Secondly, we define our own severity rating, or index of difficulty in moving from ‘amber’ to ‘green’, for each data system. A more detailed discussion of each data system follows with each classified in these ways, and finally in the overall summary and conclusions we try to indicate some possible ways forward.

### **NAO Amber Categories:**

The ‘amber’ or ‘needs improvement’ category actually comprises two sub-categories or ‘tracks’ by which improvement may be made. We have labelled these as follows:

- A Needs strengthening to ensure remaining risks are more controlled
- B Need to explain consequences of uncontrolled risks.

### **Severity ratings for this review:**

- 1 Should become green with current or planned amendments
- 2 Should become green with some further straightforward amendments
- 3 Could become green, with significant amendments
- 4 Could only become green with major amendments
- 5 Unlikely to become green in present form.

---

<sup>1</sup> Schagen, I. (2006) *Review of Data Systems Underpinning DfES SR2004 PSA Targets*, Report to DfES, March 2006.

In the brief summaries that follow for each target data system we describe the main reasons for the 'amber' category and the steps that could be taken to move to 'green', and allocate a description based on the above taxonomy.



## **Target 2(ii) - HBAI**

NAO rating: A

'Not entirely' ratings in Table 2:

- 2.1 Robust
- 2.2 Reliable
- 2.3 Comparable
- 2.4 Verifiable
- 3.1 Clear

### **Discussion**

The main risks relate to the 64% response rate to the survey and changes to the childcare question. Documentation can be improved, and the reporting could be clearer about the definition of lower income households.

### **Conclusions**

Some improvements in hand should lead to reduction of risks; however, the risk posed by the low response rate should be investigated alongside the possibilities of weighting to control it. Comparisons with LFS outcomes may allow risks to comparability to be quantified. Improved documentation and reporting should be possible relatively simply.

**Overall rating: A2**

---

## **Target 5 – Looked after children**

NAO rating: B

‘Not entirely’ ratings in Table 2:

- 1.2 Well-defined
- 2.5 Verifiable
- 3.2 Transparent
- 3.3 Comprehensive

### **Discussion**

The wording of the target can lead to misunderstandings about what is to be measured. The system is largely verifiable apart from a lack of documentation, but reporting is not entirely clear, transparent and comprehensive.

### **Conclusions**

An agreed interpretation of the target wording should be readily available, and documentation and reporting should be improved. With these changes the data system should move to ‘green’.

**Overall rating: B2**

---

## **Target 6 – KS2 attainment**

NAO rating: A

‘Not entirely’ ratings in Table 2:

- 2.2 Reliable
- 2.3 Comparable

### **Discussion**

The main risks relate to the reliability of the KS2 assessments and their marking, and to comparability over time of measures based on different tests each year. The main risk to reliability is caused by ‘borderlining’, and otherwise opinions differ on the extent to which the risk to comparability is well-controlled.

### **Conclusions**

The removal of ‘borderlining’ would go a long way towards the removal of risks to reliability, and comparability could be better assured by regular studies along the lines of the Massey report. A ‘green’ status could be achieved by further work along these lines.

**Overall rating: A3**

---

## **Target 7 – KS3 attainment**

NAO rating: A

'Not entirely' ratings in Table 2: 2.2 Reliable  
2.3 Comparable  
plus for 7(iii): 1.1 Relevant  
1.2 Well-defined  
2.4 Verifiable

### **Discussion**

The discussion for 7 (i), (ii), (iv) and (v) is identical to that for target 6. The ICT assessment {7(iii)} has greater risks, being currently based on unmoderated teacher assessments.

### **Conclusions**

Conclusions are as for target 6, except for 7(iii). Pilots of online tests for ICT are in hand, and when these are in place this should go a long way towards achieving a 'green' rating.

**Overall rating: A3/A4**

---

## **Target 8 – attendance**

NAO rating: B

‘Not entirely’ ratings in Table 2:

- 1.1 Relevant
- 1.2 Well-defined
- 2.1 Robust
- 2.2 Reliable
- 2.3 Comparable
- 2.4 Verifiable
- 3.1 Clear

### **Discussion**

This is an example where there are significant risks to the data systems, some of which might be controlled in a cost-effective fashion and others which need to be controlled by clear acknowledgement of potential shortcomings. Changes in the data system, in particular the collection of pupil-level data from 2006 (secondary) and 2007 (primary) for maintained schools will improve robustness, reliability, comparability (eventually) and verifiability. Some other changes in definitions and guidance could be made which would improve the well-defined status and perhaps reduce between-school variability.

The reporting of the target could well be improved in a way which made the remaining risks explicit and enable the data system to move towards ‘green’.

### **Conclusions**

Work in hand, plus some other changes, could well reduce many of the risks, and clear reporting of the remainder should ensure that ‘green’ status is achievable.

**Overall rating: A3 + B2**

---

## **Target 9 – take-up of sporting opportunities**

NAO rating: B

‘Not entirely’ ratings in Table 2:

- 1.2 Well-defined (‘No’)
- 2.2 Reliable
- 2.3 Comparable
- 2.4 Verifiable
- 3.1 Clear
- 3.2 Transparent
- 3.3 Comprehensive

### **Discussion**

As the PESSCL survey expands to cover all or most maintained schools, the data system for measuring quantity of provision should move towards ‘green’ status, with some caveats about comparability. However, more work needs to be done to improve the measurement of quality, based on Ofsted inspections. The reporting could well be improved in order to be clear about remaining risks to the data.

### **Conclusions**

The natural expansion of the survey to all schools should mitigate many of the risks to the measurement of quantity of provision, but it is likely that risks to the measurement of quality of provision will remain and these should be clearly reported.

**Overall rating: A2 + B3**

---

## **Target 10 - GCSEs**

NAO rating: B

'Not entirely' ratings in Table 2: 2.3 Comparable  
3.2 Transparent

### **Discussion**

The main risks to the data system are to comparability, because of the introduction of new vocational qualifications. A slight upward bias due to schools only challenging results felt to be too low is also possible. These could be controlled by clear reporting, and inconsistencies between different publications could be removed in order to move the data system towards 'green'.

### **Conclusions**

The system could become green with some fairly simple changes to reporting.

**Overall rating: B2**

---

## **Target 11 – Level 2 for age 19**

NAO rating: B

'Not entirely' ratings in Table 2: 2.4 Verifiable

### **Discussion**

The data system itself is largely fit for purpose, with the main risk being related to a need for procedures to more formalised and better documented.

### **Conclusions**

With improved documentation and a more formal set of procedures, the system could become green.

**Overall rating: B2**

---

## **Target 12 - NEET**

NAO rating: B

'Not entirely' ratings in Table 2:

- 1.1 Relevant
- 2.2 Reliable
- 2.3 Comparable
- 3.1 Clear
- 3.2 Transparent
- 3.3 Comprehensive

### **Discussion**

The data system is by and large the most cost-effective way of obtaining the required measure, although not without a number of risks. Some minor changes might reduce these, but it is the reporting of the risks which currently gives the data system an 'amber' rating. Estimates of possible sources of error could be published in addition to LFS sampling errors, and those individuals who might erroneously be included in the NEET group after the process of elimination has taken place need to be set out clearly. There is in practice a fair bit of work which can be done to improve reporting and move towards 'green' status.

### **Conclusions**

Clear discussion of the assumptions inherent in the calculation and sources of uncertainty need to be included in all reporting in order to move towards green status.

**Overall rating: B3/4**

---

## **Target 13(ii) – reducing lack of NVQ2 skills**

NAO rating: A

‘Not entirely’ ratings in Table 2:

- 2.1 Robust
- 2.2 Reliable
- 2.4 Verifiable (‘No’)
- 3.2 Transparent

### **Discussion**

There are a number of risks to this data system, many because the LFS on which it is based is not designed primarily for this purpose. 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. 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, nor is reporting transparent.

### **Conclusions**

It is not clear that the current data system could be amended in a cost-effective way to become ‘green’ in all these respects. An alternative source of data could do the job, but this is likely to be a major task. Improvements to documentation and reporting, on the other hand, should be possible which could mitigate the risks.

**Overall rating: A4/B3**

---

## **Target 14(i) – HE participation**

NAO rating: B

‘Not entirely’ ratings in Table 2: None

### **Discussion**

The main threat to verifiability is the correction for false matches when checking for prior HE experience. There are a few other details and reporting issues which can be amended fairly simply to move this to ‘green’ status.

### **Conclusions**

Some minor changes to definitions and reporting should move this to green.

**Overall rating: B2**

---

## Overall Summary and Conclusions

Target data systems which have been classified ‘amber’ are, generally speaking, those which have the potential to be fit for purpose but which need some modifications to become ‘green’ and thus suitable for the target (even if not completely ideal). NAO’s further description of each data system divides them into two categories, which we have labelled ‘A’ and ‘B’.

Category A is described as follows:

*The data system addresses the majority of risks to data quality but needs strengthening to ensure that remaining risks are adequately controlled.*

Category B has the following rubric:

*The data system addresses the majority of risks to data quality but includes limitations that cannot be cost-effectively controlled; the Department needs to explain the implications of these more clearly to the reader.*

These in essence describe two ‘tracks’ by which a data system can be reclassified as ‘green’, and in this supplementary review we have attempted to describe the possible tracks for each target data system in somewhat more detail. However, there is often a degree of uncertainty in the best track for each system. A system which contains a number of risks can in principle overcome these by either removing or controlling each one, or by documenting clearly what each is during reporting – or by a mixture of the two. In many ways the reporting option would appear simpler and require less effort – why expend resources removing a risk when clear documentation would have the same effect?

However, the NAO’s descriptions above seem to imply that risks which can be cost-effectively controlled should have this done; only when this is not possible should track B (clear explanation of risks) be brought into play. Judgements about what is and is not cost-effective need to be made, and it is not totally clear how these judgements should be made and who is to be the final arbiter. NAO’s categorisation of tracks for each data system imply that they have applied a judgement of this type in each case before allocating category A or B.

For some target systems a mixture of tracks may be most appropriate. For example, Target 8 (attendance) was classified by NAO as B, but it seems that some mitigation of risks could be achieved cost-effectively by the switch to pupil-level data collection, while the rest are controlled by reporting. Target 9 (take-up of sport), on the other hand, can have many of the risks to measuring quantity controlled by changes which are in hand, while risks to the measurement of quality probably need to be dealt with through clear reporting.

In the following table we have attempted to illustrate our conclusions for each target data system using coloured bars to show how each might move towards a ‘green’ status most effectively.

Target	Description	Track	1	2	3	4	5
1	FSP	A					
		B					
2	HBAI	A					
		B					
5	Looked after children	A					
		B					
6	KS2	A					
		B					
7	KS3	A					
		B					
7(iii)	KS3 ICT	A					
		B					
8	Attendance	A					
		B					
9	Sport	A					
		B					
10	GCSEs	A					
		B					
11	Level 2 age 19	A					
		B					
12	NEET	A					
		B					
13(ii)	NVQ2 skills	A					
		B					
14(i)	HE participation	A					
		B					