Reception baseline assessment quality assurance

Quality Assurance Solutions Tribal Education Ltd: report on findings from pilot

April 2017

Contents

List of figures

List of tables

Executive Summary

Funded reception baseline assessment took place for the first time in England during the first half term of the 2015 academic year.

The reception baseline is a new form of assessment and, in July 2015, the Standards and Testing Agency (STA) commissioned Tribal Education's Quality Assurance Solutions (QAS) team to develop and trial several approaches to quality assuring the reception baseline assessments.

This report describes the pilot that was undertaken during September and October 2015 with a sample of 150 schools. The pilot provided a mechanism to gauge the effectiveness of different methods of quality assuring the baseline assessment process to ensure that data from the assessments was valid and reliable. This was important given the potential variation in the timing of assessments, the ages of children in the reception year, the lack of written evidence (because there is no written evidence produced by the pupil), the potential for schools to reflect lower results and the speed with which young children learn (and therefore the need to capture what they know at the start of schooling, before specific teaching programmes begin.)

The **Introduction** to the report provides the background, context and overall purpose of the pilot.

The **Methodology** section explains the design and development of the 3 Quality Assurance (QA) Models that were piloted with the schools, specifically how the QA criteria were developed, and how the Quality Assurance Professionals (QAPs) were selected, trained and deployed. It also describes the process of identifying the pilot schools in a way that ensured the QA Models were tested equally against the 3 different approved reception baseline assessment schemes. It provides a description of the evidence collection tools (documentation, report templates, questionnaires and surveys) that were developed prior to the QA activities taking place.

Analyses of all of the data collected are presented in the **Findings** section which includes analysis of the extent to which the tools and activities featured in each of the 3 QA Models enabled reliable and robust evidence to be collected in relation to each of the QA criteria.

The findings presented are based on quantitative evidence obtained through analysis of the data that were collected, as well as qualitative evidence in the form of feedback from the QAPs undertaking the visits, along with feedback from the schools about their experience of the quality assurance process.

In summary, Model 1 (pre-visit questionnaire plus observation visit with advance notice) costs the same as Model 3 and provides the best opportunities for collecting robust evidence from more than one source. Model 3 (unannounced visit plus post-visit questionnaire) provides few opportunities to gather and triangulate evidence but enables observers to see the authentic assessment process. Model 2 (teleconference interview plus short observation visit) is slightly cheaper than the other models, and is the most difficult to organise and much less effective than Model 1 in terms of evidence collection.

The views of the 7 QAPs are also reported in the Findings section. Four felt that Model 1 was the most successful in terms of quality assuring the baseline assessment process to ensure that data from the assessments was valid and reliable, 2 considered Model 3 the most successful and 1 preferred Model 2. Four found Model 1 the easiest to set up and 3 thought Model 3 was easiest. Five considered Model 2 the hardest to set up whereas the remaining 2 found Model 3 the most difficult.

Schools were also surveyed about their experience of the QA visit. Overall, in response to the question asked about their satisfaction in the way in which the visit was conducted by the QAP, the result was 100% positive.

As well as the activities themselves, findings are presented in relation to the timing of the baseline assessments compared with the timing of the QA visits. This turned out to be critical as, where assessments couldn't be observed, in the majority of cases (81%) this was due to assessments having been completed before the visit.

In order to inform a consideration of the costs involved in a full roll out of each of the QA models, the amount of time and effort that each one required has been tracked and reported.

Finally, the report draws **Conclusions** and makes **Recommendations** for any future roll out of a process to quality assure reception baseline assessment. Although Model 1 was more successful at providing evidence than the other 2 models, the recommendations take into account advantages and disadvantages across all 3 models. Factors to take into account when costing QA visits and related activities are considered and recommendations are made concerning the qualifications and experience required for QAPs as well as the importance of sufficient training, support and monitoring.

1. Introduction

In September 2015 the Department for Education (DfE) introduced the reception baseline assessment:

"… to support the accountability framework and help assess school effectiveness by providing a score for each child at the start of reception which reflects their attainment against a pre-determined content domain and which will be used as the basis for an accountability measure of the relative progress of a cohort of children through primary school…[1]"

The reception baseline assessment is linked to the learning and development requirements of the Early Years Foundation Stage (EYFS) and key stage 1 (KS1) national curriculum in English and mathematics. In order to provide as level a playing field as possible, schools need to administer the assessment within the first half term of a pupil's entry to a reception class.

The following 3 schemes were approved for use in schools on 1 July 2015 by STA:

- BASE provided by the Centre for Evaluation and Monitoring, Durham University (CEM)
- EExBA provided by Early Excellence (EEx-BA-R)
- Reception Baseline Assessment provided by the National Foundation for Educational Research (NFER)

Schools were free to choose which scheme they felt best fit their needs and approach to assessment. Although the approved schemes each assess the same core areas, they vary significantly in their approach:

- BASE is teacher/teaching assistant-led assessment that is computer based for the core areas
- EEx-BA-R is a practitioner-led assessment which gathers evidence through a range of means, including observation
- The NFER scheme is a resource-based assessment with a mixture of tasks and observational checklists

In order to provide an accurate accountability measure, the data from the reception baseline assessment would need to be valid and reliable and, although scheme providers are required to quality assure the data that is submitted by the schools, STA recognised that this would need to be supplemented by a degree of central QA.

Working in collaboration with STA, Tribal Education's QAS team:

- devised a methodology that would elicit both quantitative and qualitative data to inform DfE's approach to quality assuring (monitoring) baseline assessment in

---

[1] www.gov.uk/guidance/reception-baseline-approval-process-for-assessments

schools and would be effective across all 3 of the approved baseline assessment schemes

- developed documentation to be used by QAPs undertaking QA visits
- provided training for QAPs and monitoring of QAP activities to ensure consistency and reliability
- identified and recorded any difficulties involved in undertaking QA activities
- considered the effectiveness of the model in 'detecting inappropriate behaviour'
- explored the cost effectiveness of each of 3 QA models in relation to cost
- assessed the level of expertise required by the QAP in order to make recommendations regarding expertise and qualifications required by QAPs to undertake QA visits to schools

The following sections of this report provide detail about the activities developed and implemented for the pilot study and outline the key findings, providing recommendations about approaches to quality assurance that could be implemented in a full roll out.

2   Methodology

This section provides the description and rationale of the selected approach to each of the following stages of the Reception Baseline Assessments QA project:

- Developing and selecting the QA models
- Selection and training of QAPs
- Selecting schools for the pilot
- Developing evidence collection tools and documentation
- Analysing evidence to provide key findings
- Making recommendations.

2.1   Developing and selecting QA models

'QA model' is the terminology that was adopted to refer to the different approaches to quality assuring the reception baseline assessment schemes. Initially QAS developed 5 potential QA models from which STA selected the 3 that were piloted during September and October 2015.

The development of the 5 initial models was done using a workshop approach. The workshop team included:

- an EYFS specialist
- 2 former primary headteachers
- a current early years practitioner
- 2 Tribal senior managers with experience of Ofsted inspections and the design and development of a range of quality review frameworks

The starting point was to articulate a shared understanding of the purpose of QA in relation to the reception baseline assessment, which was summarised as:

"The overall purpose of the QA process is to provide evidence that the Baseline Assessment produces outcomes and judgements that are reliable, accurate and consistent. Ideally, every pupil should be assessed against the same 'barometer' as any other pupil being assessed in any reception class across the country…[2]"

Prior to the workshop, each member of the team undertook self-study in order to familiarise themselves with the attributes of each of the 3 baseline assessment schemes to ensure that the proposed QA criteria and approaches could be applied across the different schemes.

---

[2] Reception Baseline Assessments Quality Assurance (QA): Selecting QA Models, Tribal Quality Assurance Solutions, 29 July 2015

During the workshop, 20 unique criteria were identified and these were grouped into 5 broad areas:

1. Leadership and management
2. Physical environment
3. Emotional environment
4. Administering the assessments
5. Handling data

For each of the 20 criteria the workshop participants considered the various possible QA activities that could be employed, including:

- face-to-face interviews with staff and others
- telephone interviews
- observations
- questionnaires (pre and/or post-visit)
- surveys

The QA activities were grouped to form the 5 potential models:

- Model 1: one week's notice with pre-visit questionnaire
- Model 2: one week's notice with post-visit questionnaire
- Model 3: 24 hours' notice
- Model 4: telephone interview followed by short visit
- Model 5: unannounced visit and follow-up questionnaire

Once the 5 models had been documented, QAS representatives met with STA's baseline assessment team to agree the final 3 models. During this process, some of the criteria were broken down and the final framework covered 39 criteria in 4 broad areas:

- Leadership and management (12 criteria)
- Enabling environment (14 criteria)
- Administering the assessment (11 criteria)
- Handling data (2 criteria)

The 5 potential models were further refined in order to arrive at the final 3 models and to agree for each of those models:

- the form that the QA visit would take, including the proposed length of the visit
- the level of expertise required by the representative undertaking the QA visit
- the activities that would be undertaken during the QA visit
- an outline of the documentation that would be completed by the representative undertaking the QA visit
- the potential issues likely to be detected through the proposed model and those unlikely to be detected

Table 1 Overview of 3 QA models

Overview of 3 QA models	Model 1	Model 2	Model 3
Up to 10 days' notice of visit	✓		
Pre-visit questionnaire	✓		
Teleconference interview followed by short observation visit		✓	
Unannounced visit			✓
Post-visit questionnaire			✓

2.2 Selecting and training the QAPs

Careful selection and training of the QAPs was paramount to the validity of the outcomes from the pilot. Tribal has an extensive network of educational associates with experience of EYFS and also of carrying out Ofsted type inspections and the QAPs were selected from this pool taking into account:

- **geographical location** – the pilot needed to cover a varied demographic across different parts of the country
- **qualifications and experience** – in order to gain the trust of the schools being visited, the QAPs needed to have held a senior position in a school, and/or have extensive teaching experience and/or be a qualified inspector
- **availability to attend the training** – the training was mandatory
- **ability to commit 3 days a week for 7 weeks –** as there was a requirement for all baseline assessments to be completed by half-term each QAP would have to undertake an average of 5 visits per week

7 QAPs were selected for training (plus a reserve who undertook the training so there would be an additional QAP as contingency for sickness or other unforeseeable absence).

There were 2 lead QAPs working on the project who trained, assessed and observed the QAPs. Both of the lead QAPs had extensive early years' experience at a senior level as well as experience of delivering training.

Prior to the training event participants were provided with details of, and links to, information on the selected reception baseline assessment schemes.

The face-to-face training event included training in relation to:

- an overview of the 3 baseline assessment schemes
- an introduction to the 3 selected QA Models and consideration of how each would be applied during visits
- introduction to the documentation and submission
- identification of protocols, including professional conduct and relationships with schools
- information assurance and confidentiality
- undertaking relevant QA activities (for example observation and interviewing techniques)
- procedures in the case of unexpected events such as teacher absence or inconsistent administration of assessments
- practice in completion of written documentation

The format of the training was interactive, with participants reviewing baseline assessment scenarios and then completing the associated documentation.

At the end of the training event, the Lead QAPs assessed each of the QAPs to ensure that they were ready to undertake QA visits. Following the training, and to ensure consistency in approach, the Lead QAPs accompanied each QAP on one visit in order to quality assure their work, following which, QAPs were authorised to continue to carry out the pilots.

Following Tribal's usual vetting procedures each QAP was issued with an identity card and a secure, encrypted Tribal laptop.

2.3 Selecting schools for the pilot

The size of the sample of schools was determined by what was considered the optimum size of the QAP team, together with the 6-week window in which the baseline assessment had to take place.

There was no attempt to cover every part of the country. Instead, schools were selected with a geographical spread but in clusters in order to make the process more efficient. 20 different local authorities were included across 6 regions.

Figure 1 Regions in pilot

Regions

West Midlands 13%
London 14%
North East 17%
South West 28%
South East 15%
North West 13%
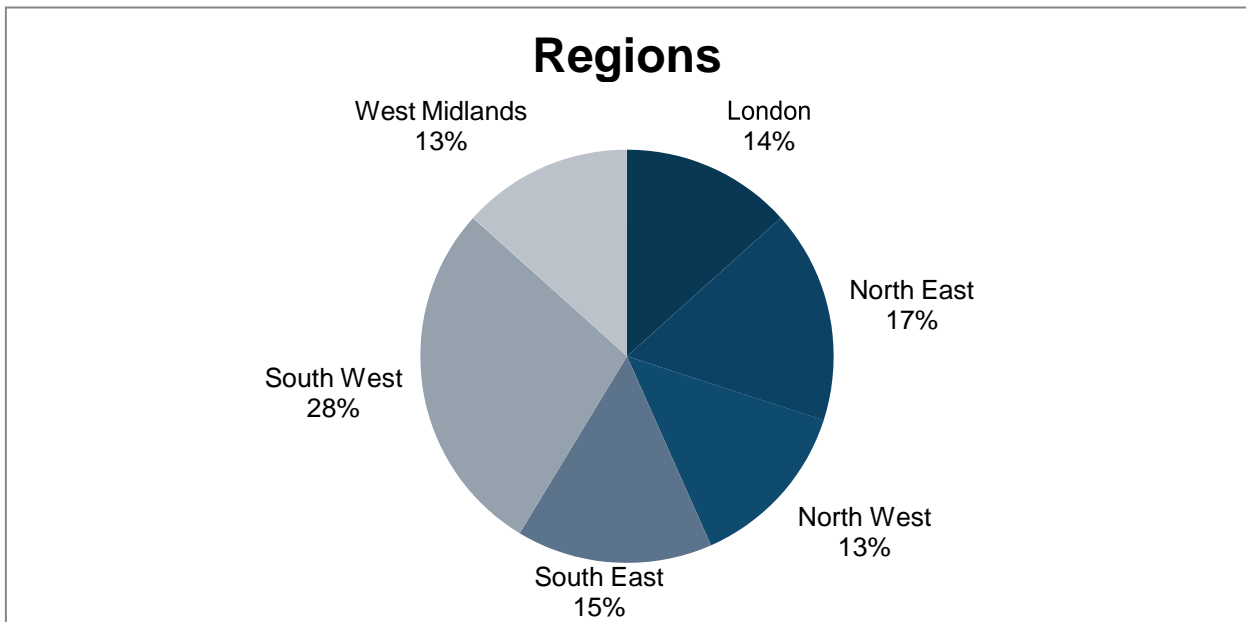
150 schools were selected that were broadly representative based on attainment, school type and size. Schools were divided into attainment bands based on their current KS1 to key stage 2 (KS2) progress measures and each band had a minimum of 10 schools within the pilot cohort.

Figure 2 indicates the broad mixture of schools within the pilot with varying attainment band ratings.

Figure 2 Schools by attainment band

Number of schools by attainment band

Band	1	2	3	4	5	6	7	8	9	10	No rating
Number of Schools	11	15	16	21	11	13	13	15	13	18	4
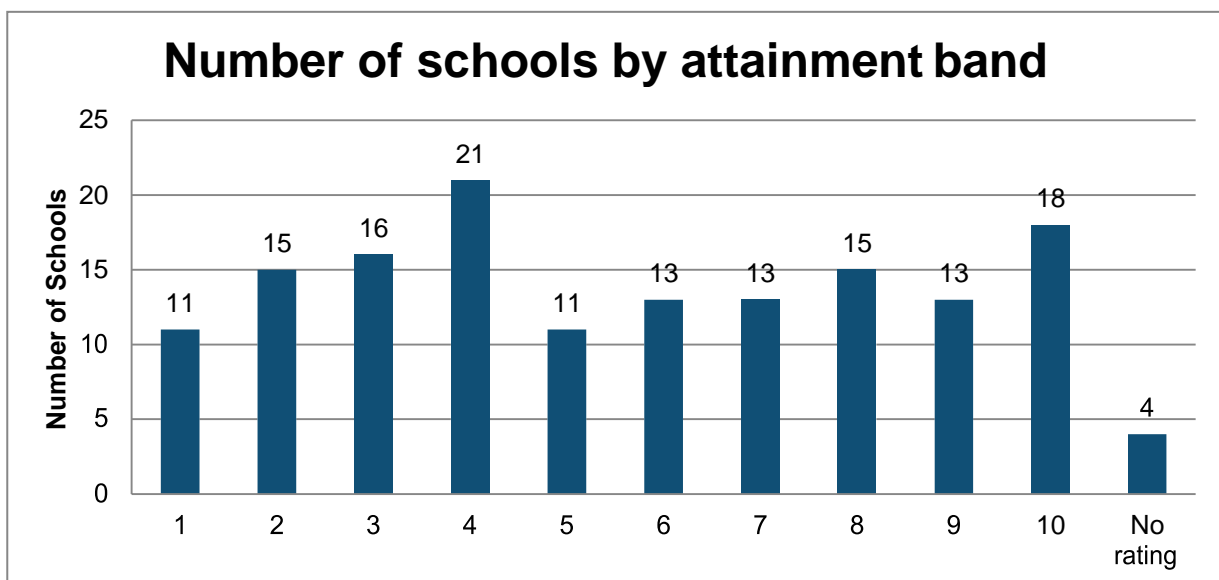
Schools were also chosen according to which of the baseline assessment schemes they had selected so that the pilot would include a fairly equal distribution of schemes.

Figure 3 Distribution of Baseline Assessment Schemes

Distribution of Baseline Assessment Schemes

NFER
32%

CEM
33%

EExBA
35%
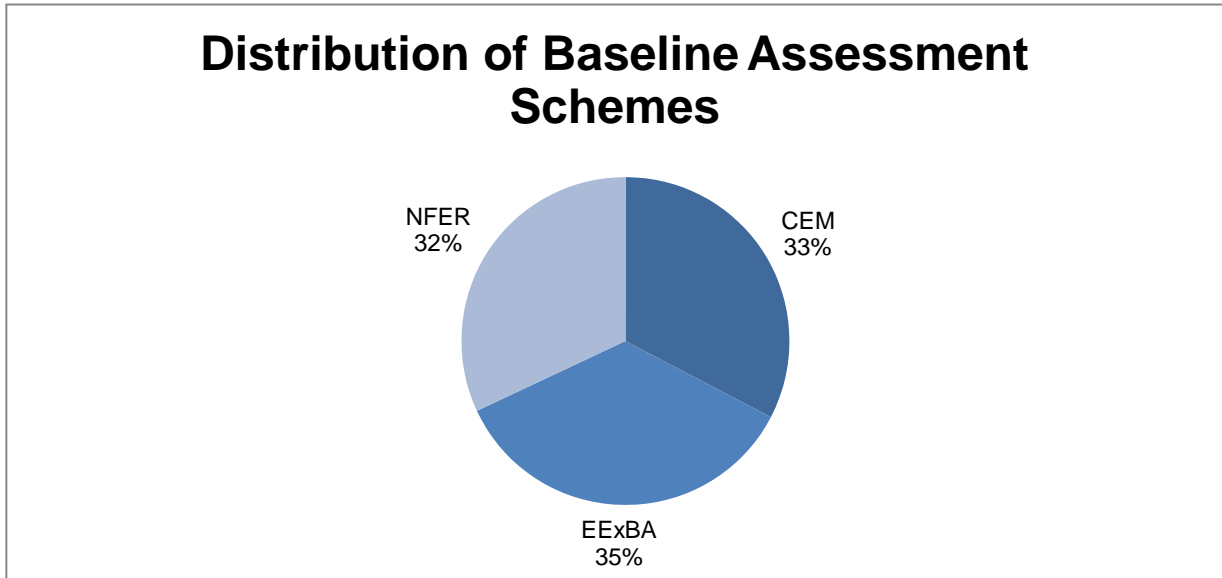
Having identified the schools each school was allocated 1 of the 3 QA models. Assignment of models to baseline assessment schemes was also as evenly balanced as possible. Each QAP was assigned a schedule of schools that ensured equal distribution across QA models and baseline assessment schemes.

Figure 4 Distribution of QA models

QA models across schools

Model 3
33%
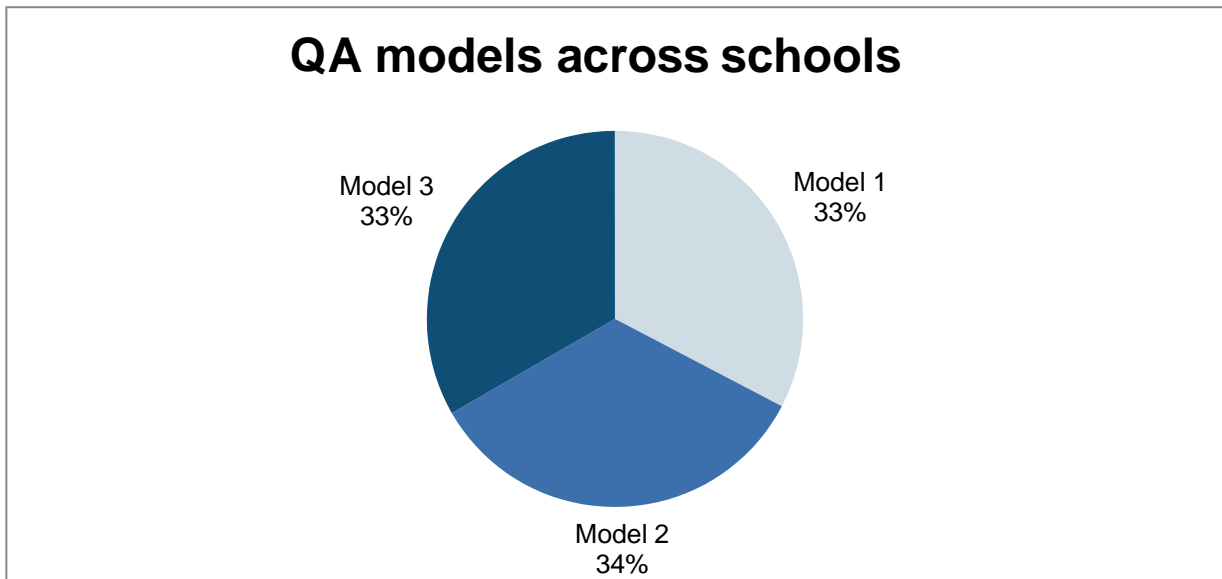
Model 1
33%

Model 2
34%

Figure 5 Distribution of Baseline Schemes per QA model
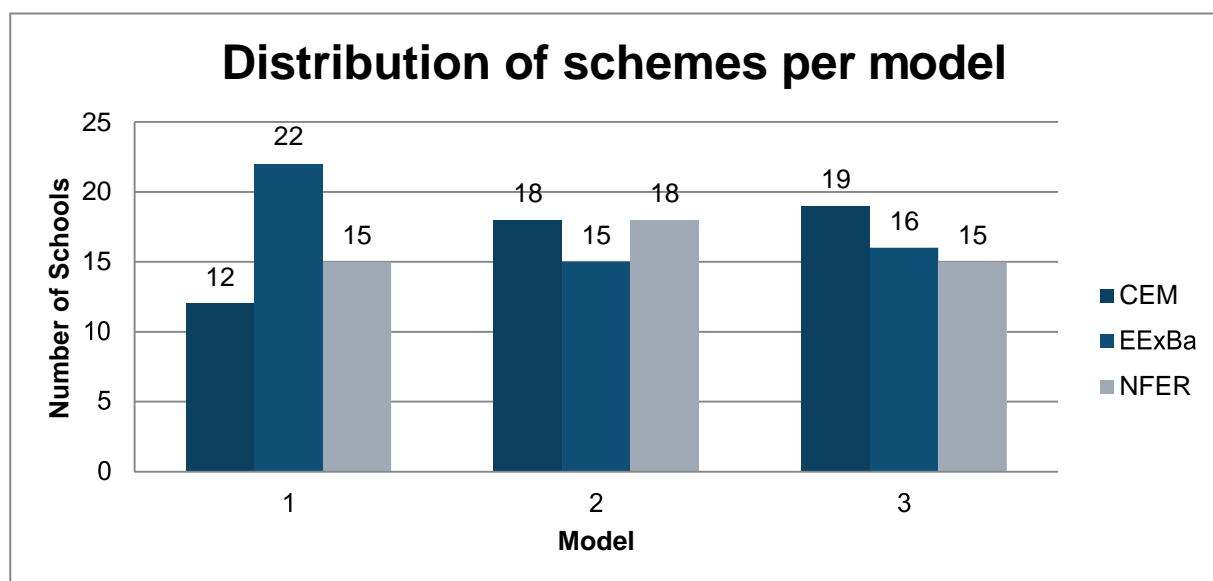
STA had written to every school prior to the start of the pilot explaining the purpose and alerting them to the possibility of an unannounced visit.

Although there were 150 schools to be included in the quality assurance pilot, 8 schools declared they did not want to be involved in the process as it was not mandatory. 2 of these schools were undergoing an Ofsted inspection at the time so wanted to concentrate their efforts on that instead. A further 4 schools did not respond to the telephone contact from the team when visits were being set up, despite contact being attempted at least 3 times. 7 of the schools declined access and of these 6 were from the unannounced Model 3 cohort. Table 2 shows the breakdown of reasons why schools did not take part.

Table 2 Reasons for not taking part

Reason	Number of Schools
Did not want to be involved	8
No response	4
No access to school	7
Total	**19**

A further group of schools were selected to replace those where QAPs had been unable to engage with schools in the original sample. Therefore, in total, 150 schools were included in the quality assurance pilot.

2.4 Collecting evidence

The methodology for collecting evidence was based on the premise that, where possible, there should be more than 1 source (often referred to as triangulation). All evidence for each school (however it was collected) was summarised by the QAP on a 'smart' PDF report template. This enabled the QAP to complete the report offline and then submit it online.

The report template provided a balance of yes/no answers, Likert scale answers and free text responses. This produced a range of both quantitative and qualitative data.

The report enabled QAPs to provide their own observations and professional judgements about the extent to which each of the models had allowed them to quality assure the reception baseline scheme being operated in the particular school. The success rate of this measure was dependent on the professional judgements of QAPs who indicated whether the model had been 'very successful', 'successful', 'partially successful' or 'unsuccessful' in allowing them to quality assure the process.

Immediately following the process, schools were invited to provide feedback about their experiences of the QA event. This follow-up survey sought responses in 4 areas:

1. Overall, satisfaction with the way the visit was conducted by the QAP
2. Whether the input required by respondents and other school staff was reasonable
3. Whether the questions that were asked (during any interview and on any questionnaire if one was completed) were relevant to determining the quality of the baseline assessment process
4. Whether, during the visit, the QAP engaged appropriately with school staff and children

The QAPs were invited to respond in a structured debriefing teleconference with a Lead QAP and the Project manager about their perceptions of the process. The questions focused on the difficulties involved in undertaking the visit, the effectiveness of the different models in quality assuring the baseline assessment process, use of reporting formats and the QAPs' professional judgements and reflections regarding future quality assurance procedures.

Table 3 provides a description of each piece of documentation that was used to collect evidence.

Table 3 Documentation to collect evidence

	Description of QA documentation to be completed by QAP and School	Completed by QAP	Completed by School
QA report template (all models)	This will provide the main source of data to be analysed for the final report. There is a single report template for all 3 QA Models, and the same report format will be used for each of the Baseline Assessment Schemes. The report template has been designed to ensure that the QAPs record against all of the criteria consistently. Many of the responses are input as yes/no, tick boxes and Likert-type scales which will enable comparative analysis to take place. There is also scope for QAPs to provide additional comments. The QAP will use the interview and observation pro-formas as well as questionnaires completed by the school to populate the report and to provide additional evidence. When completing the report, QAPs will highlight the sources of evidence for each of the criteria covered.	✓	
Model 1 pre-visit questionnaire	Very brief (2 page) questionnaire to help schools think about the visit and to provide information on management and leadership of baseline assessment.		✓
Model 1 pro-forma	This combines the interview template with the observation template, so that the QAP just has 1 document per visit. It covers everything that was not covered in the pre-visit questionnaire.	✓	
Model 2 pro-forma	Model 2 does not include a questionnaire, so the telephone/teleconference interview pro-forma covers all of the criteria. Again, 1 pro-forma covers both interview and short observation visit.	✓	
Model 3 pro-forma	This is the unannounced visit model, so the interview and observation pro-forma covers all criteria, with the expectation that not all of them will be covered during the visit. It will also provide a record of any visit where neither interview nor observation was possible.	✓	
Model 3 post-visit questionnaire	This questionnaire is designed as a table with un-numbered rows. This will allow the QAP to amend (tailor) the questionnaire following the visit by deleting any of the criteria that were		✓

	Description of QA documentation to be completed by QAP and School	Completed by QAP	Completed by School
	already covered during the visit.		
Survey for Schools	Following the visit (or other activity where a visit did not take place) each school will be emailed or sent a short survey to ascertain their level of satisfaction with the process.		✓
Final feedback form QAPs	At the end of the pilot, each QAP provides feedback on the process overall, including their professional views on the relative merit of the different models, issues arising.	✓	

After the first few visits and a review of the documentation received from the QAPs, 'calibration' teleconferences were held with the QAPs to ensure that the evidence collection tools were working as intended. This involved making some small adjustments to the documentation for the remainder of the pilot. These were only concerned with ease of collection and therefore had no impact on the outcomes of the pilot.

2.5   Analysing evidence

The report, which was completed by the QAP for each school, formed the primary source of evidence. As an interactive PDF form it fed directly into a database of all responses, which allowed Tribal's data team to provide quantitative analyses in a number of areas including:

- percentage of schools where observations did not actually take place (and the reasons why)
- the effect of the timing of QA visits on the quality of outcomes
- when schools were undertaking assessments
- observations of QAPs about the level of consistency across assessors, reliability and accuracy of assessments
- the relative ease of setting up and organising the 3 QA models
- satisfaction levels of schools for each of the models

Additionally, the free text responses allowed the QAPs to comment on the suitability of whichever QA model was employed when quality assuring the particular assessment scheme in use at the school.

The feedback from the schools was analysed to provide evidence of the relative success of the 3 models.

2.6 Drawing conclusions and making recommendations

The conclusions and recommendations in this report are made on the basis of a sample of 150 schools having QA activities undertaken by 7 QAPs trained and supervised by experienced Lead QAPs. Detailed written evidence was provided in a consistent format in relation to each individual school, QA model and baseline assessment scheme.

The level of detail and range of this evidence should be sufficient to provide confidence in the recommendations made in relation to the advantages and disadvantages of each of the 3 models in the pilot, specifically in relation to:

- the potential costs of a full roll out each of the models across the country
- any difficulties with the perceived models
- the extent to which each model is helpful in detecting 'inappropriate' behaviours or other factors that might cause the baseline assessment to be unreliable

It is important to note that the purpose of this pilot was not to provide feedback to STA or the schools about the quality of the assessment processes that individual schools are implementing. However, in order to test each of the 3 QA models, the QAPs were asking questions and making observations as if they actually were quality assuring for these purposes.

3 Findings

The pilot has provided a significant amount of quantitative and qualitative evidence in relation to the advantages and disadvantages of each of the 3 models.

This section provides analyses and evaluations of the evidence that underpins the conclusions and recommendations made in the final section of the report, including:

- a quantitative analysis of the extent to which each QA Model enabled the QAP to collect reliable and robust evidence in relation to each of the 39 QA criteria
- evaluation of both quantitative and qualitative evidence concerning the direct observations that took place
- an examination of when, during the half-term, the baseline assessments were undertaken
- details concerning the timing of the visits and related QA activities
- the amount of time that was required to undertake the activities for each of the 3 QA models
- a review of the quantitative and qualitative evidence received in the form of feedback from the schools in relation to each of the QA models
- further quantitative and qualitative evidence received from the QAPs during the end of pilot debriefing session that was held with each of them
- a review of the qualifications and experience of the QAPs

3.1 Reliability of evidence in relation to the QA criteria

Throughout the pilot evidence was collected against 39 criteria in the following 4 areas:

1. Leadership and management
2. Enabling environment
3. Administering the assessment
4. Handling data

In order for a QA model to provide a framework that collects robust and reliable evidence it is desirable if that evidence comes from more than 1 source. When completing the reports, QAPs had to document the sources of evidence against their response for each of the criteria.

This section of the report provides an analysis against the QA criteria for each of the models in relation to whether there is:

- no evidence
- 1 type of evidence
- more than 1 type of evidence

3.1.1 Leadership and management

Table 4 lists the criteria relating to leadership and management aspects of the QA process.

Table 4 Leadership and management criteria

1.	Leadership and management
1.1	Is there an identified leader in the school for Baseline Assessment?
1.2	What contribution has the leader made to ensure consistency of administration (for example development of a plan; leadership of training)?
1.3	How many staff are involved in administering assessments?
1.4	How have they been selected?
1.5	What, if any, training have they received?
1.6	How useful was the training?
1.7	What procedures has the school put in place to ensure the assessments are administered consistently and that the outcomes are reliable and accurate (including arrangements for internal moderation)?
1.8	What use have you made of any support for technical or content related queries available from the scheme provider to staff administering the assessments?
1.9	How effective was this support?
1.10	When are the assessments scheduled to take place?
1.11	In what order, if any, have children been identified to be assessed (for example older children first or children who attended the school nursery or other)?
1.12	On what basis have you made those decisions?

Table 5 shows the extent to which each of the 3 models was able to detect elements of leadership and management of the baseline assessments in a school.

Table 5 Evidence sources for 'Leadership and Management'

QA Criteria Reference	QA Model 1 (planned visit)			QA Model 2 (phone call and follow up visit)			QA Model 3 (unannounced visit)		
	No Evidence	1 type of evidence	More than 1 type of evidence	No Evidence	1 type of evidence	More than 1 type of evidence	No Evidence	1 type of evidence	More than 1 type of evidence
1.1; 1.2	2%	41%	57%	0%	84%	16%	4%	88%	8%
1.5; 1.6	2%	45%	53%	2%	90%	8%	6%	92%	2%
1.7	4%	61%	35%	2%	90%	8%	2%	96%	2%
1.8; 1.9	4%	76%	20%	6%	90%	4%	2%	98%	0%
*1.10	8%	55%	37%	2%	94%	4%	8%	90%	2%
1.11;1.12	4%	53%	43%	0%	96%	4%	4%	92%	4%
Average	**4%**	**55%**	**41%**	**2%**	**91%**	**7%**	**4%**	**93%**	**3%**

Model 1 was able to provide more than 1 type of evidence against the leadership and management criteria on a significantly greater number of occasions than Models 2 and 3.

(* QA criterion 1.10 relates to the question "when are the assessments scheduled" and where there was no evidence this was sometimes due to that question not being answered on the post-visit questionnaire.)

3.1.2 Enabling environment

Table 6 lists the criteria relating to the enabling environmental aspects of the QA process.

Table 6 Enabling environment criteria

2.	Enabling environment
2.1	Where will the assessments take place?
2.2	What factors affected this decision?
2.3	Has the environment had any impact on the assessment process and, if so, how?
2.4	Does the assessment accurately capture what the child is able to do with particular reference to children with identified special needs or disabilities, including those with sensory, attentional or physical issues?
2.5	In relation to children for whom English is an additional language, to what extent does the assessment accurately capture what the child is able to do?
2.6	Have any of the augmented or additional adaptations, recommended for SEND by providers, been used to improve accessibility?
2.7	Were there any children for whom the assessment was difficult to manage?
2.8	How, if at all, has the assessment procedure been introduced to children and their parents?
2.9	Have parents raised any concerns around baseline assessment? Has there been any parental pressure (on the child or on the school)?
2.10	What is the relationship of the assessor to the child (for example this may be: class teacher, key person, class teaching assistant, HLTA or other)?
2.11	How long, on average have the majority of children known the assessor (for example since September 2015, September 2014 or other).
2.12	Have any of the results been unexpected to the school?
2.13	If a child appeared unable to respond or to be unclear about expectations or procedures, or was anxious, upset or worried during the assessment process what actions would the assessor take (for example stop the assessment, take a break from the assessment, re-start the assessment later)? What proportion of children actually became upset or distressed?
2.14	Is the nature of interaction/administration between assessor and child consistent across all assessors?

Table 7 shows the extent to which each of the 3 models was able to detect relevant attributes of the environment in which the baseline assessments were carried out.

Table 7 Evidence sources for 'Enabling Environment'

QA Criteria Reference	QA Model 1 (planned visit)			QA Model 2 (phone call and follow up visit			QA Model 3 (unannounced visit)		
	No Evidence	1 type of evidence	More than 1 type of evidence	No Evidence	1 type of evidence	More than 1 type of evidence	No Evidence	1 type of evidence	More than 1 type of evidence
2.1; 2.2	2%	61%	37%	2%	75%	24%	4%	84%	12%
2.3	12%	65%	22%	0%	82%	18%	4%	92%	4%
2.4	8%	82%	10%	2%	92%	6%	2%	98%	0%
2.5	16%	80%	4%	14%	84%	2%	14%	86%	0%
2.6	10%	90%	0%	8%	90%	2%	16%	84%	0%
2.7	4%	86%	10%	4%	88%	8%	2%	96%	2%
2.8	6%	80%	14%	0%	88%	12%	2%	94%	4%
2.9	12%	88%	0%	8%	92%	0%	4%	96%	0%
2.1	2%	47%	51%	2%	88%	10%	6%	82%	12%
2.11	6%	94%	0%	2%	94%	4%	0%	98%	2%
2.12	6%	86%	8%	6%	86%	8%	6%	94%	0%
2.13	6%	82%	12%	4%	90%	6%	8%	90%	2%
2.14	20%	59%	20%	14%	71%	16%	12%	78%	10%
Average	**9%**	**77%**	**15%**	**5%**	**86%**	**9%**	**6%**	**90%**	**4%**

Model 1 was able to provide more than 1 type of evidence against the criteria associated with an enabling environment on more occasions than Models 2 and 3. However, Model 1 appears to provide no evidence on more occasions than Models 2 and 3. Closer scrutiny of this data shows that in some cases the QAPs selected "no evidence" when in fact the question may not have been relevant at the particular school. This should be considered in any revision of the report template so there is an option that differentiates between "not applicable" and "no evidence provided".

QA criterion 2.5 refers to the question "in relation to children, for whom English is an additional language (EAL), to what extent does the assessment accurately capture what the child is able to do?" There is a relatively high "no evidence" response to this question across the board and this reflects the proportion of schools in the pilot without EAL pupils in their classes. Again, a "not applicable" option would have made this clearer.

3.1.3 Administering the assessment

Table 8 lists the criteria relating to administering the assessments.

Table 8 Administering the assessment criteria

3.	Administering the assessment
3.1	Have entry arrangements (such as a staggered entry/induction period) influenced the timing of assessments?
3.2	How has the school chosen to structure the assessments? (For example, each of the core areas assessed in turn for all children or all the assessments for one child completed in a set period.) Has the structure had an impact?
3.3	How practical is it to follow precisely the administrative guidance provided by the Scheme provider?
3.4	If, for any (technical) reason, an assessment was not completed in the required way, what procedures would be followed? (Technical reason could be computer failure or, for example, a component missing from the assessment materials.)
3.5	To what extent is it possible to take into account information about the child's learning from multiple sources, such as parents or the previous provider?
3.6	When recording outcomes how are a child's unique needs taken into account?
3.7	Are assessors able to benchmark judgements about a child's learning so that there is internal consensus about whether a child has or has not achieved a particular learning point? Is this the case for a) all 3 core areas b) some core areas c) other?
3.8	If an assessor believed the outcome of an assessment to be inaccurate, what action, if any, would be taken?
3.9	Did the school feel there was sufficient time to prepare for the assessment?
3.10	What, if anything, has the school learned that would help others to implement the assessment more effectively.
3.11	Having administered it this time, would you have done anything differently next time in preparation?

Table 9 shows the extent to which each of the 3 models was able to detect how the assessment is administered in the school.

Table 9 Evidence sources for 'Administering the Assessment'

QA Criteria Reference	QA Model 1 (planned visit)			QA Model 2 (phone call and follow up visit)			QA Model 3 (unannounced visit)		
	No Evidence	1 type of evidence	More than 1 type of evidence	No Evidence	1 type of evidence	More than 1 type of evidence	No Evidence	1 type of evidence	More than 1 type of evidence
3.1	10%	86%	4%	2%	92%	6%	0%	98%	2%
3.2	4%	84%	12%	4%	80%	16%	6%	90%	4%
3.3	2%	94%	4%	6%	90%	4%	4%	94%	2%
3.4	24%	73%	2%	14%	82%	4%	20%	80%	0%
3.5	10%	86%	4%	10%	80%	10%	6%	94%	0%
3.6	18%	80%	2%	12%	82%	6%	10%	88%	2%
3.7	4%	90%	6%	4%	88%	8%	2%	98%	0%
3.8	16%	82%	2%	6%	88%	6%	6%	94%	0%
3.9	8%	88%	4%	6%	92%	2%	2%	98%	0%
3.10	10%	86%	4%	10%	82%	8%	16%	84%	0%
3.11	14%	84%	2%	8%	86%	6%	12%	88%	0%
Average	**11%**	**85%**	**4%**	**7%**	**86%**	**7%**	**8%**	**91%**	**1%**

The instance of 'more than one type of evidence' relating to the administration is very low across all models with the highest being Model 2 with a 7% average, followed by Model 1 with a 4% average and then Model 3 with 1%. This is because, in the case of these criteria, the only way for the QAPs to get the information was by asking the school.

QA criterion 3.4 had a particularly high "no evidence" response rate. This question is "if, for a technical reason, an assessment was not completed in the required way, what procedures would be followed?" In the majority of cases this question had not been/could not be answered by the schools. This is where a 'not applicable' option on the report template would help differentiate the responses and allow further analysis.

3.1.4 Handling data

Table 10 lists the criteria relating to data handling aspects of the process.

Table 10 Handling data criteria

4.	Handling data
4.1	How are outcomes from assessments recorded? a) Manually during the assessment then transferred to the online recording system b) On-line during the assessment
4.2	Are checks and balances in place to ensure that any data that is inputted after or during the assessment process is accurate?

Table 11 shows the extent to which each of the models was able to detect how assessment data is handled by the school.

Table 11 Evidence sources for 'Data Handling'

QA Criteria Reference	QA Model 1 (planned visit)			QA Model 2 (phone call and follow up visit			QA Model 3 (unannounced visit)		
	No Evidence	1 type of evidence	More than 1 type of evidence	No Evidence	1 type of evidence	More than 1 type of evidence	No Evidence	1 type of evidence	More than 1 type of evidence
4.1	10%	53%	37%	10%	69%	22%	8%	88%	4%
4.2	4%	80%	16%	6%	84%	10%	6%	92%	2%
Average	**7%**	**66%**	**27%**	**8%**	**76%**	**16%**	**7%**	**90%**	**3%**

Model 1 was able to provide more than 1 type of evidence against the handling of assessment data criteria on a significantly greater number of occasions than Models 2 and 3.
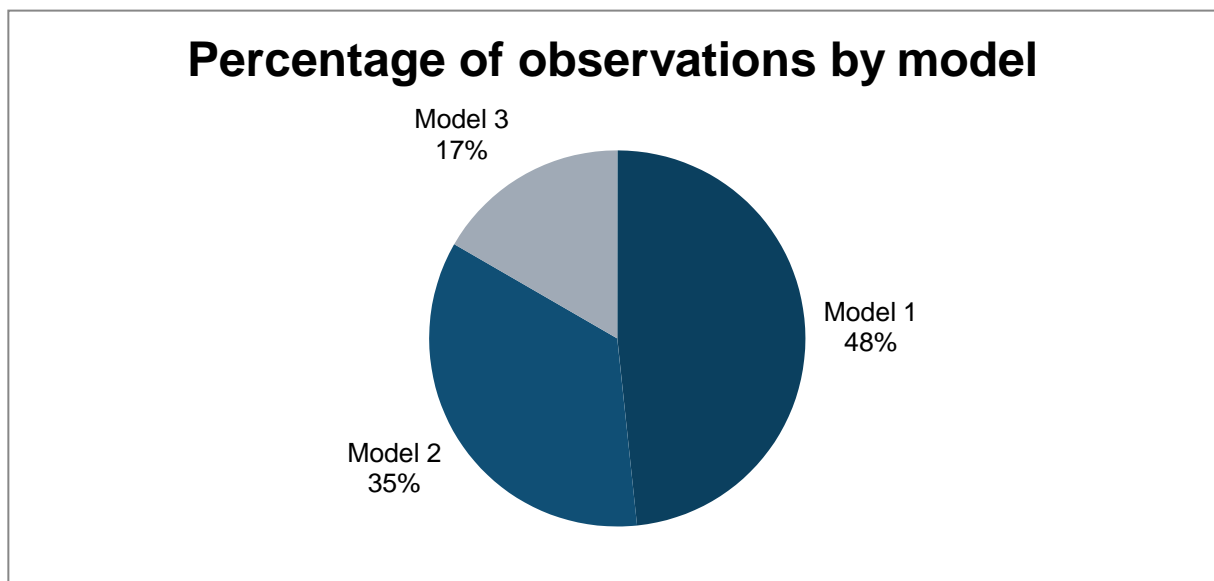
3.2 Observations of assessment

Each model provided an opportunity for the QAP to observe baseline assessments taking place, however this was completely dependent on whether schools had already completed the assessments or if they were scheduled at the time of the visit. In order to maximise the effectiveness of the QA visit, completion of both discussion and observation elements of the QA process was required. The completion of both elements varied considerably depending on the model.

### 3.2.1 Number of observations taking place per QA Model

Figure 6 shows the breakdown of all of the observations that took place throughout the pilot. Almost half (48%) of the observations undertaken were from the Model 1 visits. 35% of the observations related to Model 2 and only 17% of the observations related to Model 3. This is due to the unannounced nature of Model 3 where the school did not know the visit was happening and therefore could not plan to accommodate the observing of assessments.

Figure 6 Percentage of observations per model

The reason assessments were not able to be observed was mainly (81%) due to assessments having already been completed. Figure 7 shows the breakdown of reasons why observations of assessments were not possible at the time of the QA visit.

Figure 7 Why observation was not possible
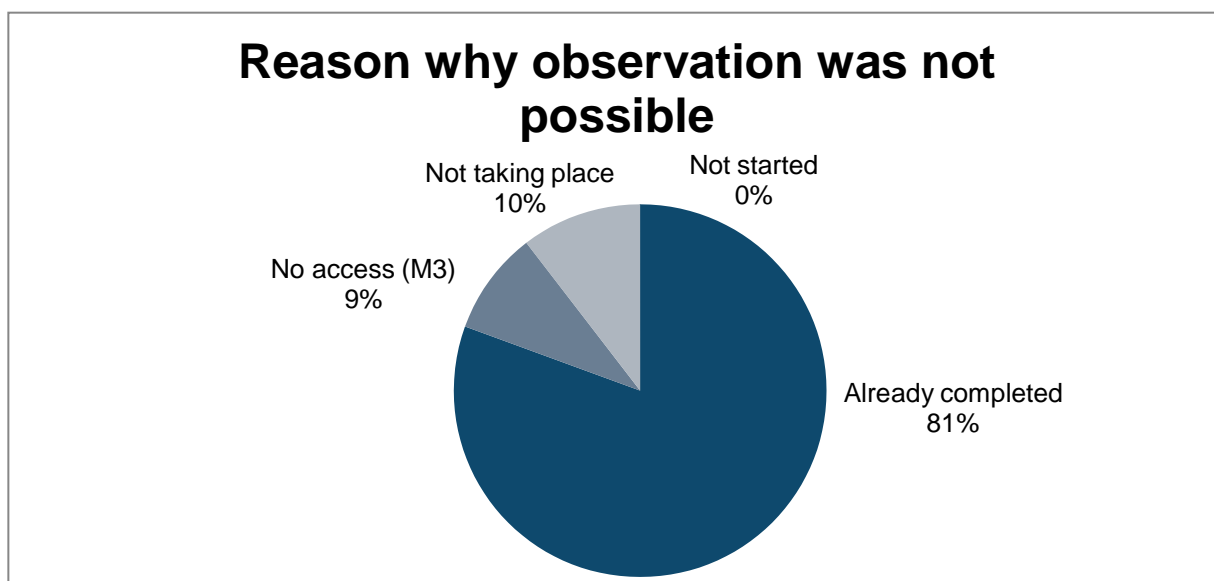
3.2.2 QAPs' comments regarding observations

The QA report included 'free text' questions where QAPs were asked to comment on the observations that took place in each school. This is a representative sample of the comments received in relation to each model:

Model 1

"Highly supportive school that had arranged for both the headteacher and EYFS lead to be available. Shared information without prompting, offering transparency and detail. QA Model was robust with this scheme and school"

"The pre-visit information, together with a full-ranging discussion during the school conversation enabled me to gain a full picture of how the school was implementing the scheme and the moderation arrangements it had made. The observation of assessment enabled corroboration of information given and also provoked further conversations which confirmed the consistency of assessment by the teacher and TA"

"This Quality Assurance Model was successful in enabling me to quality assure the implementation of the baseline scheme. The school had sufficient time to plan my visit and arrange for me to observe some assessment activities. The school completed the pre-visit questionnaire well in advance of the visit"

Model 2

"This model worked well. I had enough time to see the assessments taking place and to see the consistency across the staff"

"Because I was able to speak to the FS class teacher on the phone and in person during the visit, I gained a good picture of her assessment plan and its implementation..."

"This Quality Assurance Model was only partially successful as there were no observations made of the scheme being administered…"

Model 3

"Although the headteacher was not in the school that day, the EYFS lead was very welcoming and offered for me to observe an assessment plus gave me some time to have a professional discussion…"

"Upon arrival I met the headteacher and deputy headteacher. Due to staffing levels that day, it was not possible to complete an interview…"

"This visit was highly successful. The headteacher was very accommodating, and took over the reception class in order to release the teacher…"
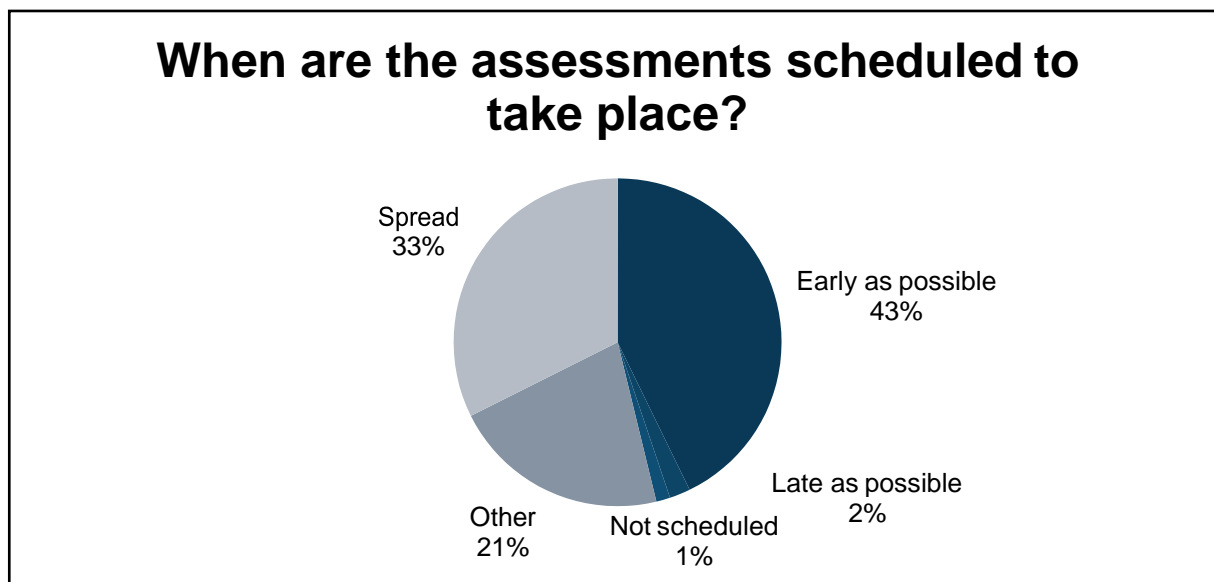
3.3   Timing of assessments and QA visits

3.3.1 Timing of the assessments

43% of the schools in the pilot explained that they planned to carry out the baseline assessments as early in the term as possible, with almost half of those being schools who were conducting the CEM assessment scheme. The majority of schools conducting EExBA-R assessments (57%) planned to spread their assessments across the term.

One of the QAP team commented that assessments were deemed more useful in the morning than afternoon by some schools as children get tired by the afternoon.

Figure 8 indicates when in the half term the assessments were scheduled to take place.

Figure 8 Timing of assessments

3.3.2 Timing of QA visits

The reception baseline assessments are administered within the first half term of a pupil's entry to a reception class. This means there is a maximum of 8 weeks for schools to conduct the baseline assessments and for any QA observations to take place. In order to allow schools the first week of term to welcome children and organise the baseline assessments, there were no QA visits conducted in this week as part of the pilot. The first QA visit took place in the week commencing 14 September 2015 with the final visit occurring in the week commencing 26 October 2015.

Figure 9 shows the declining percentage of visits where assessments were able to be observed by week. It should be noted that week 8 only had 3 visits, all of which were able to include an observation, hence the skewed result of 100% in the final week.

Figure 9 Percentage of visits per week that included an observation

Percentage of visits per week that included an observation
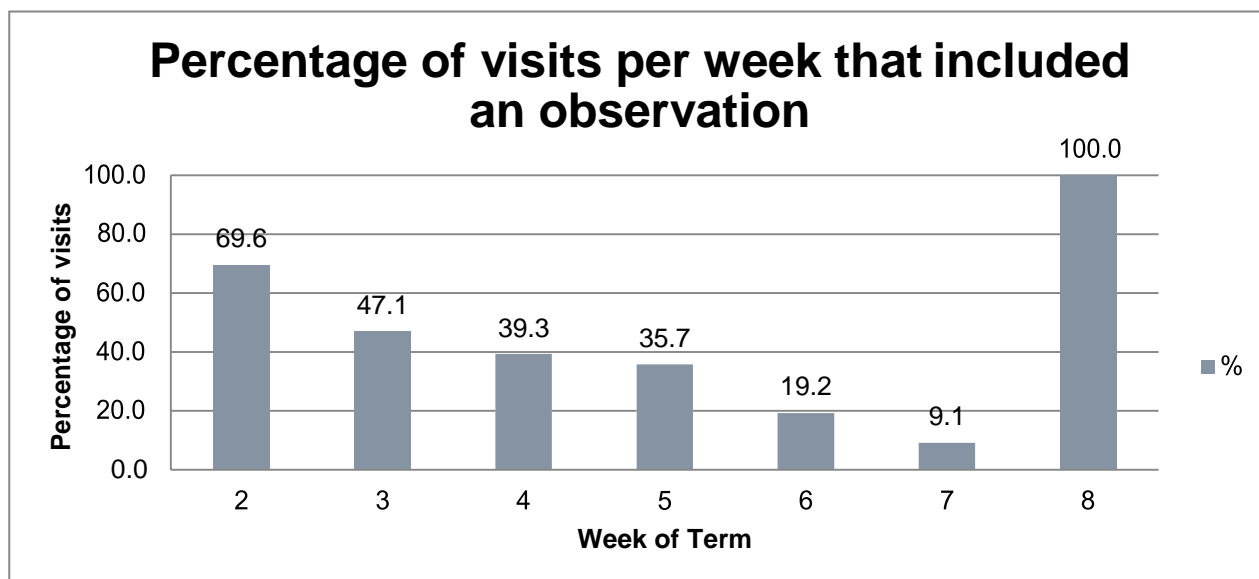
3.4 Time taken to carry out activities for each QA model

Table 12 indicates the average length of time for each of the models to be carried out in terms of the visit and where appropriate the teleconference as well.

Table 12 Time taken to carry out activities for each QA model

	Average time for visit	Average time for teleconference	Average time combined
Model 1	1 hour 34 minutes	N/A	1 hour 34 minutes
Model 2	47 minutes	43 minutes	1 hour 30 minutes
Model 3 (where visit went ahead	1 hour 20 minutes	N/A	1 hour 20 minutes
Model 3 (where QAP was refused entry)	11 minutes	N/A	11 minutes

The Model 3 timings have been split to show how long on average was spent when the QAP was given access to the school (1 hour and 20 minutes) and also for the occasions where a QAP arrived at the school but was subsequently turned away (11 minutes).

This pilot indicates that on average the two key elements of the Model 1 QA process took the longest but only by 4 minutes compared with Model 2. Model 3 visits where access was allowed took the least time. It is important to note that none of these times include travel or the time it took to organise the activities associated with the different QA models. These are discussed in the section "organising the QA visit".

3.5   Feedback from schools

It was recognised that the feedback from schools would be an important way to judge their perception of the effectiveness of the QA models. The return rate for the survey was 55 of the 150 schools (37%) which approximates to slightly more than one third of participating schools.

Schools were asked to respond to the following questions:

- Overall, I am satisfied with the way the quality assurance visit was conducted by the Quality Assurance Professional (QAP)
- The input required by me and other school staff was reasonable
- The questions that were asked (during any interview and on any questionnaire if one was completed) were relevant to determining the quality of the baseline assessment process
- During the visit the QAP engaged appropriately with school staff and children

Schools were then asked to provide one of the following ratings against each question:

- Strongly agree
- Agree
- Don't know
- Disagree
- Strongly disagree

3.5.1 Satisfaction rating

When asked if the school was satisfied with the way the QA visit was conducted by the QAP, the result was 100% positive. 71% of respondents indicated they strongly agreed and 29% agreed. Comments from schools included:

"The way the whole process was undertaken provided the opportunity to engage in a professional discussion about the process"

"I found the visit to be very professional and the dialogue that took place was interesting and helped us reflect on our procedures for administering the baseline assessments"

3.5.2 Input required by school

96% of respondents either strongly agreed (54%) or agreed (42%) that their input to the process was reasonable. One school disagreed and one school strongly disagreed. The school which strongly disagreed was part of the Model 3 QA process which is unannounced and stated:

"An unannounced visit to QA an assessment that is not statutory was not appropriate. It took me away from an important meeting and disrupted the provision in the foundation stage. If this Model is to be used it needs to be in line with the practice for monitoring the administration of phonics screening or SATs. I.e. to take place whilst the baseline is being administered not at the end of a school day"

The school that disagreed was also part of the unannounced Model 3 cohort and stated:

"The input required of me and the other staff was reasonable had we had warning. As we had a no notice check it was difficult as staff had to be released with no notice and without cover staff to put in place. A short notice period would have been preferable"

3.5.3 Relevant questioning

58% of respondents strongly agreed and a further 38% agreed with the statement *"The questions that were asked (during any interview and on any questionnaire if one was completed) were relevant to determining the quality of the Baseline Assessment process"*. There were 2 schools who indicated they did not know and there were no schools that disagreed or strongly disagreed. One Model 2 school commented about the telephone interview:

"I held my session via phone call after the school day. I enjoyed the chance to share our school findings and how we have introduced and managed the baseline. It was nice to feel like I had a voice and felt very listened to - a pleasant experience all round"

3.5.4 QAP engagement at school

98% of respondents either strongly agreed (76%) or agreed (22%) that the QAP engaged appropriately with school staff and children. The remaining school indicated they did not know as the QAP had only liaised with them over the phone and had not visited the school because the assessments had already been completed and therefore there was no requirement for a visit.

Comments from the surveys included:

"This was a positive experience; the visit was professional and purposeful"

"I would highly recommend all assessors to be from an Early Years background"

"The amount of notice provided for the visit was appreciated as opposed to a no notice visit"

"I am so pleased that the visit was conducted by an Early Years specialist. She was completely sympathetic to the requirements of the assessment and how it could challenge individuals who may not adhere to the 'norm' because of additional needs"

"It was a very professional and worthwhile experience. The QAP was very knowledgeable and encouraged professional dialogue that put the validity and worth of the baseline at the heart of the conversation. We valued her experience in the EYFS and this, along with her non-intrusive and professional manner, contributed to an informative and worthwhile process for both parties"

"…they need to see tests in progress!"

3.6 Feedback from QAPs

There were 7 QAPs, each of whom took part in a debriefing session and this section outlines the responses to questions about each of the 3 QA models.

3.6.1 Set up of QA model

Figure 10 shows that when asked *"which, if any, type of QA visit was easier to set up?"* the majority of the team (4 QAPs) felt that Model 1 was easiest to set up, followed by Model 3 (3 QAPs) and none of the team felt that the Model 2 QA visit was easy to set up.
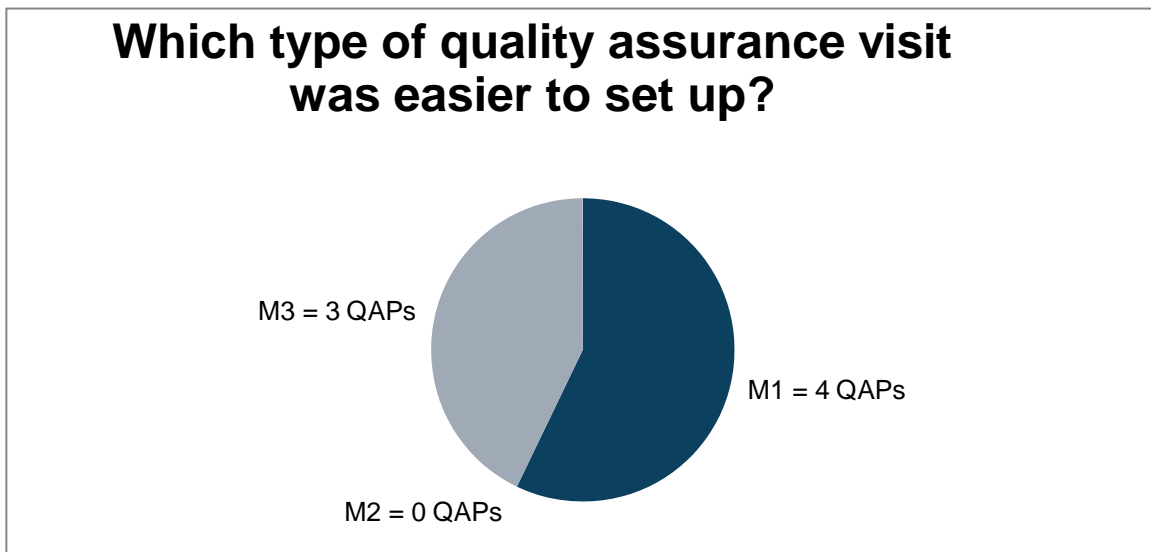
Figure 10 Set up of QA model

Figure 11 shows that when asked *"which, if any, type of QA visit was more difficult to set up?"* the majority of the team (5 QAPs) felt that Model 2 was more difficult to set up, followed by Model 3 (2 QAPs) and none of the team felt that the Model 1 QA visit was difficult to set up.

Figure 11 Model most difficult to set up

Which type of quality assurance visit was more difficult to set up?

M1 = 0 QAPs

M3 = 2 QAPs

M2 = 5 QAPs
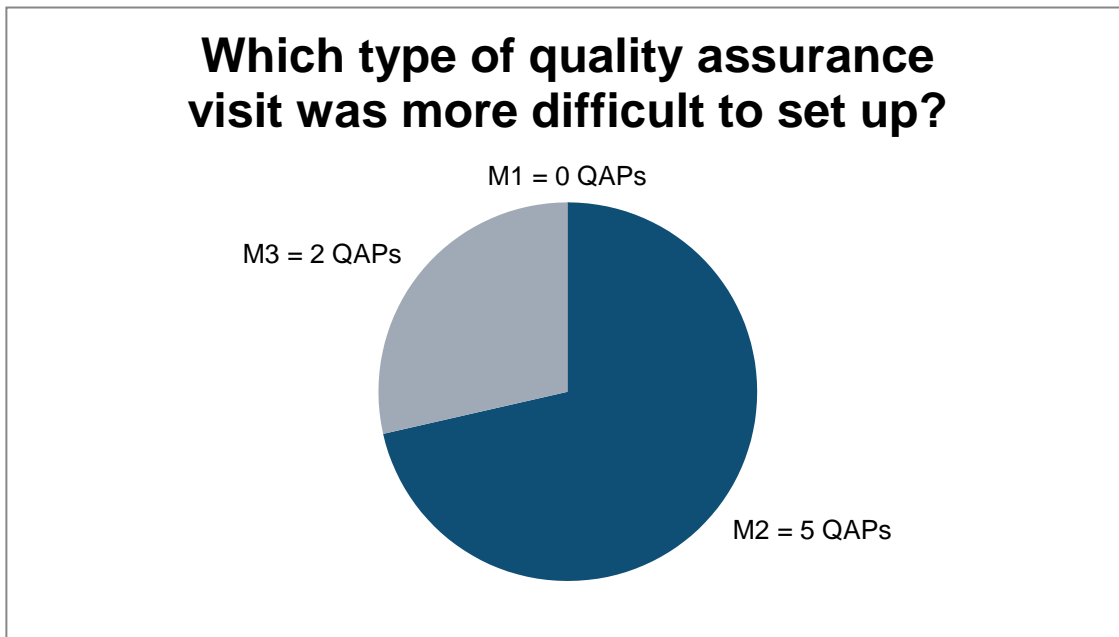
Generally, the team found that Model 1 was easiest to set up, allowed schools time to prepare for the visit and also the visit could potentially be scheduled when the assessments were taking place in order to allow an observation. However, the further into the term the harder this was to achieve as a lot of schools had already completed assessments for the later visits. The team also commented that Model 1 tended to be received better by schools because as it was pre-arranged, it was perceived to be an official visit.

Some members of the team thought that, in the case of Model 2, it was more difficult to build the rapport with the school as the first interaction and questioning were over the telephone. It was also difficult to organise suitable diary times when the appropriate people could be available for a telephone conversation and sometimes a significant number of calls was required to organise all elements. Members of the QAP team also considered that the teleconference element of Model 2 did not allow for them to pick up on important non-verbal communications such as body language, nodding and gesturing.

In general, the QAP team felt that Model 2 was neither an economical nor efficient way of conducting the QA due to the difficulties inherent in the teleconference element.

3.6.2 Outcomes and judgements

Figure 12 shows that when asked *"which of the models was most successful in ensuring that the Baseline Assessment produces outcomes and judgements that are reliable, accurate and consistent?"* 4 of the QAP team selected Model 1, 2 considered Model 3 most successful and 1 opted for Model 2.

Figure 12 Most successful QA model

Which of the models was most successful in ensuring that 'the Baseline Assessment produces outcomes and judgements that are reliable, accurate and consistent'?
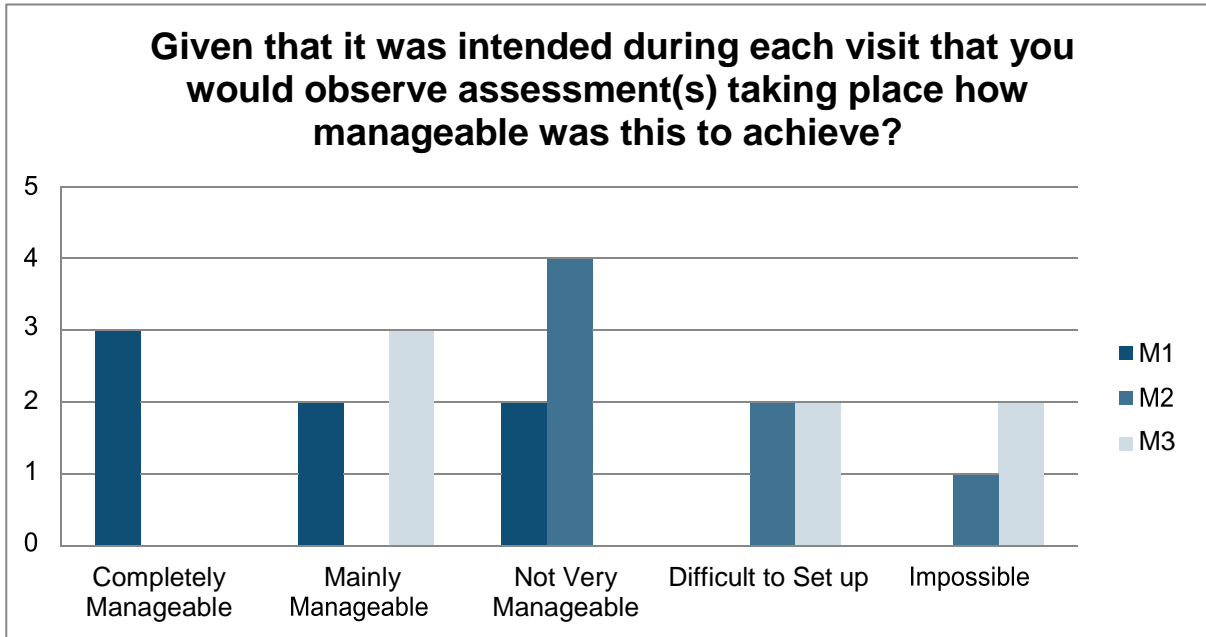
M3 = 2 QAPs

M1 = 4 QAPs

M2 = 1 QAP
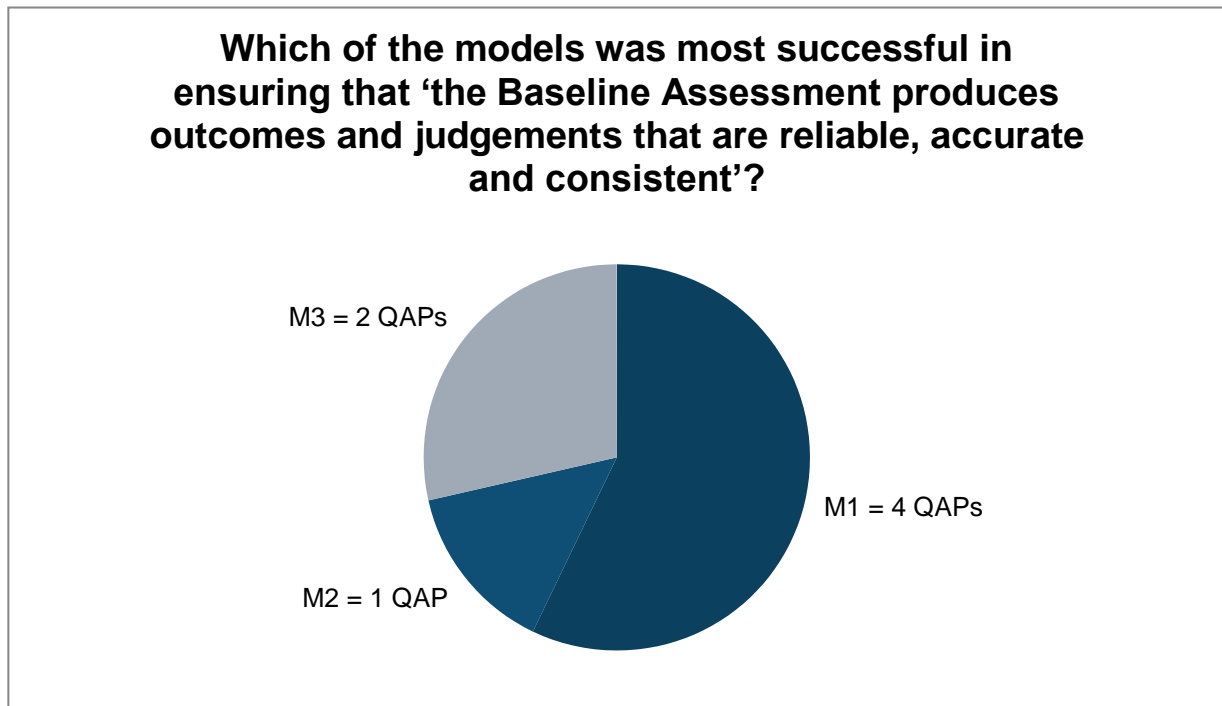
3.6.3 Observing assessments

The QAPs were also asked *"given that it was intended during each visit that you would observe assessment(s) taking place how manageable was this to achieve in relation to each of the following?"* The responses they were able to choose from were:

- completely manageable
- mainly manageable
- not very manageable
- difficult to set up
- impossible

Figure 13 shows the responses to this question against each model, with Model 1 having the majority of positive responses (completely manageable or mainly manageable). Model 2 received an almost even amount of responses in terms of mainly manageable, difficult or impossible and Model 3 had mostly negative responses (not very manageable, difficult or impossible).

Figure 13 Managing to set up an observation

Given that it was intended during each visit that you would observe assessment(s) taking place how manageable was this to achieve?

(Bar chart. Vertical axis from 0 to 5. Series: M1 (dark blue), M2 (medium blue), M3 (light blue))

- Completely Manageable: M1 = 3
- Mainly Manageable: M1 = 2, M3 = 3
- Not Very Manageable: M1 = 2, M2 = 4
- Difficult to Set up: M2 = 2, M3 = 2
- Impossible: M2 = 1, M3 = 2

It was very difficult to observe assessments in Model 3 because the visits were unannounced, so whether or not an assessment was taking place was a matter of chance.

The QAP team were also asked *"do you believe it is possible to quality assure the Baseline Assessment process successfully if the QAP is unable to observe any assessments taking place?"* Figure 14 shows the majority of the team believed this was not possible.

Figure 14 Success of QA without observation

Do you believe it is possible to successfully quality assure the Baseline Assessment process if the QAP is unable to observe any assessments taking place?

(Pie chart)
- 14% — Yes = 1 QAP
- 86% — No = 6 QAPs
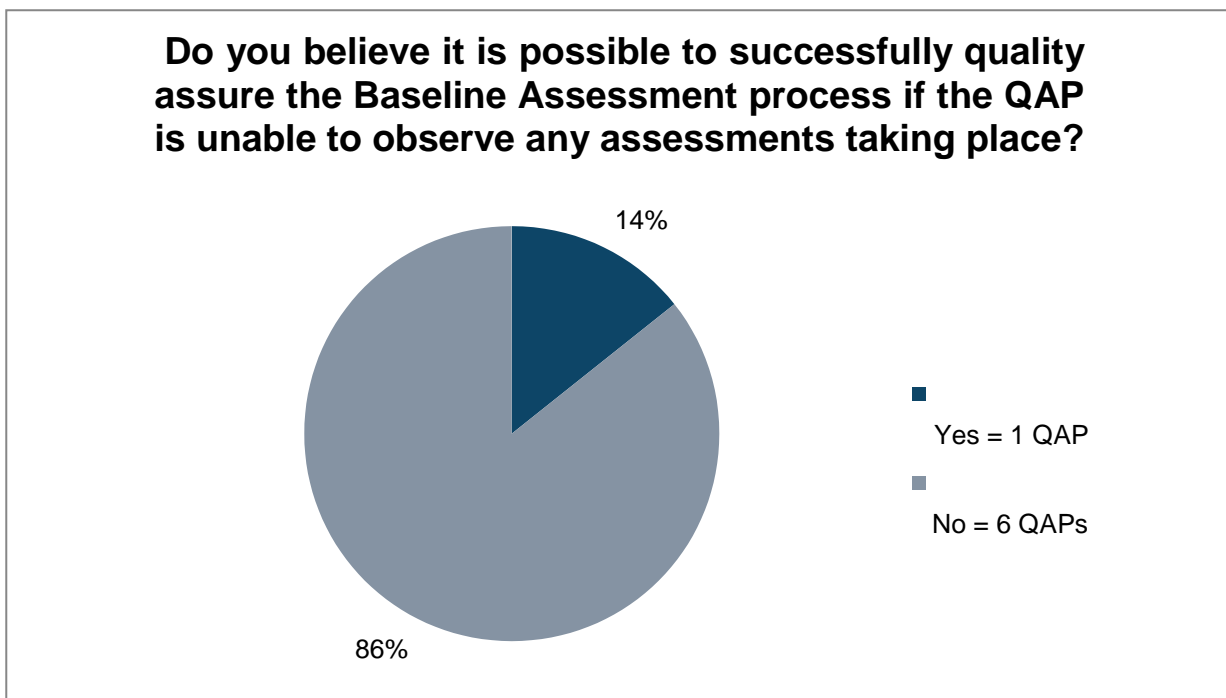
Only 1 of the team answered yes, with the remainder indicating that it was important to observe assessments being undertaken to ensure the entire process is fully quality assured. However, the team did comment that this could vary depending on which baseline assessment scheme is being observed. It was believed that observations would definitely be needed for the CEM and NFER schemes; however for those schools undertaking the EEx-BA-R scheme this may not be absolutely essential if the QAP can see observational evidence of the assessments.

It was also suggested that if there was effective, external moderation then observing the assessments may not be absolutely critical, but internal moderation would not be sufficient to provide robust quality assurance.

Finally, the team were asked if the documentation was easy to use and understand for each of the models. The team agreed that it was very simple and straightforward to use.

4. Conclusions and recommendations

This final section of the report draws conclusions and provides recommendations concerning the most appropriate approach to QA that might be implemented from September 2016. These conclusions and recommendations are based on the range of qualitative and quantitative evidence that was set out in Section 3.

Section 4.1 below outlines the advantages and disadvantages of each model through an evaluation of:

- the disadvantages and difficulties involved in undertaking the QA visit and other activities
- any advantages in the model
- the effectiveness in detecting inappropriate behaviour

Section 4.2 provides an analysis of the potential costs associated with each of the 3 models.

The advantages and disadvantages of each model are then summarised in **Section 4.3** with a recommendation for the most appropriate approach to quality assurance, as supported by the evidence from the pilot.

Section 4.4 provides some general recommendations (which apply regardless of what model is implemented) concerning:

- preferred qualifications and recommended training for QAPs
- providing information to schools
- timing of the quality assurance process

4.1 Disadvantages and advantages of each model

Table 13 Advantages and disadvantages of each model

	Model 1 (Pre-visit questionnaire plus observation visit with advance notice)	**Model 2** (Teleconference interview plus short observation visit)	**Model 3** (Unannounced visit plus post-visit questionnaire)
Advantages and benefits	• Allows the greatest number of observations to be made • Is easiest to set up • Provides the highest levels of consistency and reliability because it allows the collection of evidence from more than 1 source on a greater number of occasions than the other models • Notice of the visit is given to schools, therefore schools are able to prepare fully and organise for the appropriate staff to be available for interview • Allows for initial information to be submitted by the school in the form of the pre-visit questionnaire enabling the QAP to prepare effectively for the visit	• Notice of the visit is given to schools, therefore the school is able to prepare fully and organise for the appropriate staff to be available • No QAPs were refused access when undertaking a Model 2 QA visit	• Takes the least time to undertake the entire process (1 hour and 20 minutes) as there are no pre-visit activities involved and, if no assessments are taking place, then no observations can take place, and there may not be anyone available to interview

	Model 1 (Pre-visit questionnaire plus observation visit with advance notice)	**Model 2** (Teleconference interview plus short observation visit)	**Model 3** (Unannounced visit plus post-visit questionnaire)
Disadvantages and difficulties	• On average takes slightly longer for the entire QA process to be completed, at 4 minutes longer than Model 2 and 14 minutes longer than Model 3 • The school visit takes on average twice as long as Model 2	• The majority (71%) of the QAP team felt this model was the most difficult to set up and required a significant number of attempts to do so • The QAP team also found that the teleconference element of this model particularly challenging as they were unable to pick up on the non-verbal indicators	• Provides the fewest opportunities to gather evidence from more than 1 source • Least successful in terms of the number of observations of assessment that can be carried out • The school does not receive notification of a visit and therefore the relevant staff may not be available for interview • Least popular model from the perspective of schools based on feedback from the post-visit surveys • Most likely of all the models to result in a wasted visit, which represents poor value for money (6 out of the 7 refusals were from this model)

	Model 1 (Pre-visit questionnaire plus observation visit with advance notice)	Model 2 (Teleconference interview plus short observation visit)	Model 3 (Unannounced visit plus post-visit questionnaire)
Effectiveness in detecting inappropriate behaviour	• The school receives notification of a visit and therefore could potentially adapt their processes knowing they will be observed • Because the visit is pre-arranged, separate interviews can be scheduled with several staff, providing a range of descriptions about how the assessment is usually administered	• The school receives notification of a visit and therefore could potentially adapt their processes knowing they will be observed • The interview is by teleconference, so if several staff are on the call together, this can restrict the range of different viewpoints • The visit is a 'short observation visit' that does not include any further opportunities for discussions with staff	• The unannounced nature of the visit means the school is unable to 'stage' an assessment specifically for the observation

4.2 Costs of quality assuring reception baseline assessments

For the pilot evidence collection tools, other documentation, a help desk and associated technology were put in place and these would be required in a full roll out, regardless of the model(s) selected. Costs could be kept to a minimum by adapting pilot materials wherever possible.

An associate delivery model was utilised in the pilot because it provides scheduling flexibility and allows for the fact that the work is seasonal rather than continuous throughout the year. This worked well, enabling QAPs to plan for blocks of time and minimise travel, therefore we would recommend using associates as an efficient way of securing value for money in a national roll out.

The most significant costs comprise the effort and activities undertaken by 3 groups of people: Lead QAPs, QAPs and administrators. Drawing on the pilot experience, we have worked out the days utilised for each model and also the average number of visits that could be made by a QAP. We consider this to be 30 visits. The maximum number of QAPs in a training group is 20, therefore we have based our calculations on 20 QAPs undertaking 30 visits each, covering 600 schools. As the number of schools and QAPs increases, advantages of economies of scale can be realised and these are reflected in our calculations.

4.2.1 Lead QAPs: training and monitoring

The QAP pilot training included distance learning to enable familiarisation with the different baseline assessment schemes. Feedback from the QAPs indicates that half a day should be sufficient for this. The face-to-face training was delivered by 2 Lead QAPs, was intensive, and incorporated assessments of the participants. An important part of the training, noted by the QAPs, was the opportunity to learn through group discussion. All 3 models require the same amount of training.

For a full roll out all elements of the training would still be required. However, we consider that it would be possible to accommodate up to 20 trainees per group without removing any of the content. Ensuring that the groups contained people with differing expertise would maximise the benefits of learning through discussion.

In the pilot, all QAPs received a monitoring visit but for a roll out, consistency monitoring of 10% of the QAPs should be sufficient, given that they will have been assessed and schools will provide feedback. With careful logistical planning, two monitoring observations can be completed in one day by the Lead QAPs.

In order to engage schools with the quality assurance process and help them improve in future we would recommend that, in the roll out, schools should receive copies of the QAPs' reports. Time would need to be allocated for Lead QAPs to quality assure these documents.

Table 14 indicates the number of Lead QAP days required to train, monitor and quality assure the work of 20 QAPs, assuming that the QAPs undertake an average of 30 visits each.

Table 14 Lead QAP days (for 20 QAPs and 30 visits/reports per QAP)

Activity	Model 1	Model 2	Model 3
Trainers (2 Lead QAPs x 2 days each)	4	4	4
Consistency monitoring observations by Lead QAPs (10% of cohort, 2 observations in 1 day)	1	1	1
Quality assuring 600 QAP reports	60	60	60
Total number of Lead QAP days	**65**	**65**	**65**

4.2.2 QAP days

In addition to the days required for training, QAPs undertook the following activities as part of the pilot:

- Pre-visit
- School visit
- Post-visit

The pre-visit activities included some general research on the school such as reading the Ofsted inspection report and reviewing the website. Phone calls to set up the visit (Model 1) or the teleconference (Model 2) are also included. In the roll out, these activities would still be necessary.

The post-visit activities include preparing and sending out the post-visit questionnaire for the Model 3 schools and also completing and submitting the QA report. Again, these tasks would be required in the roll out.

Table 15 indicates the number of QAP days required to be trained and to undertake an average of 30 visits each.

Table 15 QAP days required for each QA model (assuming 20 QAPs and 30 schools each)

Activity	Model 1	Model 2	Model 3
Training for QAPs (20 x 2.5 days per QAP)	50	50	50
600 schools – pre-visit activities	60	120	30
600 school visits	300	180	300
600 schools – post visit activities	60	60	90
Total number of days	**470**	**410**	**470**

These calculations do not include any travel time or costs. These can be minimised by careful scheduling of visits by QAPs to schools that are local to them

4.2.3 Administration days

Administration days are broken down into the following:

- Report completion check and uploading to central system
- Distribution and collation of school survey
- Scheduling monitoring visits and QA reading
- General tasks

In the pilot, every report submitted needed to be checked, to ensure all areas had been fully completed by the QAP, and then uploaded to the central collation point. A survey had to be distributed to the appropriate person within each school and the results collated. Administrators scheduled monitoring visits to enable two to be conducted in one day and allocated QA reading. Other general tasks include:

- responding to QAP enquiries
- allocating schools to QAPs and reallocating when necessary
- invoice processing
- responding to any concerns that schools may have

All the above administrative tasks would be required in the full roll out. In terms of administrative time, no difference between the 3 models emerged in the pilot.

Table 16 indicates the number of administration days required to manage the process for 600 schools.

Table 16 Administration days required for each QA model (600 schools)

Activity	Model 1	Model 2	Model 3
Report uploading; checking for completion	30	30	30
School survey distribution and collation	25	25	25
Scheduling monitoring visits and QA reading	20	20	20
General tasks	20	20	20
Total Number of Administration Days	**95**	**95**	**95**

4.2.4 Total cost of models

The total number of days required to train 20 QAPs who each undertake the quality assurance of baseline assessment in an average of 30 schools (600 schools in total), to administer the whole end to end process and to monitor and assure the quality of the final reports is shown in table 17.

Table 17 Total days required for each QA model (600 schools)

Type of Days	Model 1	Model 2	Model 3
Lead QAP	65	65	65
QAP	470	410	470
Administrator	95	95	95
Total Number of Days	**630**	**570**	**630**

4.3   Recommendations for the most appropriate approach

The pilot was structured around 3 potential QA models, each made up of different approaches and evidence collection methods. The models were designed to enable the testing of different elements of quality assuring the baseline assessment, and it would have been possible to combine the elements into a different set of 3 models. Therefore, although the evidence from the pilot supports Model 1 as being the preferred model overall, it is advisable to consider combining effective elements from the other models in any further roll out.

Model 1 provided the most opportunities to elicit different types of evidence from multiple sources as it included a pre-visit questionnaire and a pre-arranged visit that would include an observation of assessment. Having received the completed questionnaire from the school prior to the visit, the QAP had the background knowledge necessary to plan interview questions and strategies for evidence collation in order to verify the information from the questionnaire.

In terms of time and effort, Model 1 took a few minutes longer than the other models, but would cost the same as Model 3 in any full roll out because much of the cost would be attributed to training, administration and management.

Where Model 1 was the weakest was in relation to its lack of effectiveness in detecting inappropriate behaviour. Because the visit was pre-arranged, it would be possible for the school to arrange for the assessment to take place in ideal circumstances, which may not necessarily reflect every day conditions. Model 3 had the potential for being the most effective in detecting inappropriate behaviour because the visit was unannounced.

However, for the same reason, it was ineffective as frequently no assessment was taking place at the time of the visit.

It is recommended that an approach be considered that includes some schools being quality assured according to Model 1 and others according to a revised Model 3. The revised Model 3 could start with a pre-visit questionnaire, with one of the questions being: *'When will reception baseline assessments be scheduled to take place in your school?'* The schools would be told to expect an unannounced visit during this period (and perhaps be given a window of two weeks in which the visit would take place).

4.4　General considerations

4.4.1 Preferred qualifications and recommended training for QAPs

The QAPs selected to undertake the QA activities had a range of experience and qualifications which is summarised in table 18.

Table 18 Qualifications and experience of QAP Team

Qualification	QAP 1	QAP 2	QAP 3	QAP 4	QAP 5	QAP 6	QAP 7
BA or equivalent		✓		✓	✓	✓	✓
Post-grad (Education)					✓		
Post-grad (Other)	✓					✓	
QTS		✓		✓	✓	✓	✓
Child care qual. (NVQ 3+)			✓				
Management qualification			✓		✓	✓	
Accredited Ofsted Inspector	✓				✓		✓
Experience							
EYFS Leader	✓	✓	✓	✓		✓	
Class teacher		✓		✓	✓	✓	✓
Headteacher		✓			✓	✓	✓
Other SMT	✓		✓	✓	✓	✓	
LA Officer/Advisor	✓	✓	✓			✓	
SEN Specialist	✓		✓				
Ed Tech Specialist				✓			
School Improvement		✓			✓	✓	✓
Governor Training/Advisor				✓			✓

All of the QAP team had experience of either being an EYFS leader or a class teacher. This enabled them to show the schools a level of understanding that was directly relevant to the school and the subject matter being quality assured. A number of schools commented in the survey that this was appreciated.

Due to the nature of the visit it was also important, in particular for those QAPs who had little or no Ofsted inspector experience, to understand fully the importance of how to approach and liaise with schools. Members of the QAP team commented this was a very useful element of the training they undertook before the visits.

The need for specialist training of those undertaking QA visits cannot be overstated. The QAPs need to understand the different assessment schemes being used and to have well developed interview and observation skills. Additionally, they need to complete the reports consistently, fully and accurately. The face-to-face training provided opportunities for QAPs to practise each of these areas and have their level of competence assessed.

4.4.2 Notification to schools

In order to maximise the effectiveness of the QA process it is important that schools receive notification at the beginning of term informing them of the QA process and potential visit (and whatever form that might take). A letter was sent out by STA to schools at the end of June for the pilot, however, a number of schools commented they did not realise they should expect a visit. One response to the schools survey indicated "*I think Heads should be notified that their school might have a visit, as it was unexpected and explanations needed to be lengthy to ensure the visitor's purpose was validated*".

4.4.3 Timing of the quality assurance process

The findings from the pilot indicate that for the QA model to be reliable, accurate and consistent, observations of assessments need to be undertaken in order to triangulate evidence that was either discussed in the teleconference or face-to-face with the staff. Model 1 was therefore deemed in general to be a more holistic approach where the QAP could discuss the assessments, see them happening and then discuss further after they had taken place. Model 2 and 3 only allowed this triangulation to occur if assessments were able to be observed, therefore the timing was critical.

For the roll out, it is important to ascertain from schools when the baseline assessments are taking place so the QA process can be scheduled accordingly. The results from the pilot indicate that the earlier in the term the QA visit takes place the higher the chance of being able to observe an assessment, in particular for those using the CEM scheme. Almost 70% of the visits in week two were able to include the observation of an assessment compared with only 9% in week 7.

4.4.4 Documentation used to carry out quality assurance

The QA documentation (report template, questionnaires and survey) used during the pilot were generally used successfully and without technical difficulty across all 3 models. It is recommended that, prior to any further roll out, the documentation be reviewed and, where necessary, revised to take into account feedback from the QAPs and issues of reporting (for example, differentiation between 'not provided' and 'not applicable').

Follow us on Twitter:
@educationgovuk

Like us on Facebook:
facebook.com/educationgovuk