An investigation into how to assess the quality of education through curriculum intent, implementation and impact

Phase 3 findings of curriculum research

This report outlines what we have done in phase 3 of our research into the quality of curriculum in schools.

Corporate member of
Plain English Campaign
Committed to clearer communication

361

Contents

An investigation into how to assess the quality of education through curriculum intent, implementation and impact
December 2018, No. 180035

2

Introduction

In January 2017, Her Majesty's Chief Inspector commissioned a major research study into the curriculum. The purpose of this research was to ensure that Ofsted could assess the quality of education in a valid and reliable way. Because a refocus on the curriculum – the substance of what is taught in schools – has been at the heart of our recent proposals for the new education inspection framework (EIF), the evidence from this study will play a prominent part in how inspectors will inspect the quality of education in the future.

We carried out the study in three distinct phases. We visited 40 schools in phase 1,[1] 23 schools in phase 2[2] and 64 schools in phase 3. In phase 1, the research attempted to understand the current state of curricular thinking in schools. It found that many schools were prioritising test and exam results and teaching a curriculum that was narrowly focused on those tests and exams instead of thinking about the substance of education – knowledge, skills and progression.

Phase 2 of our research sought to look at schools that had particularly invested in curriculum design. The main findings were three-fold:

- First, the discussion format with senior leaders around curriculum design and management was an appropriate method for identifying their curricular intent for their pupils. These discussions were rich in detail and allowed inspectors to identify whether the curriculum design had been thought about thoroughly or not by the participants. Inspectors saw it as a potential starting point for engagement on the curriculum.

- Second, despite its usefulness, it was difficult to form an overall assessment on the curriculum quality each school was providing through the discussion format alone. We agreed that additional methods for assessing curriculum implementation, particularly as a means of testing the accuracy of leaders' thought on their curriculum design, were needed.

- Third, the discussions revealed a number of common themes on the curriculum that might be important to assess in greater detail. Although we went to schools with very different approaches to the curriculum, we did find some common factors that appear related to curriculum quality. These included:
  - the importance of mapping subjects as individual disciplines
  - using the curriculum to address disadvantage and provide equality of opportunity

---

[1] HMCI's commentary: recent primary and secondary curriculum research. Ofsted, October 2017; www.gov.uk/government/speeches/hmcis-commentary-october-2017.

[2] HMCI's commentary: curriculum and the new education inspection framework, Ofsted, September 2018; www.gov.uk/government/speeches/hmci-commentary-curriculum-and-the-new-education-inspection-framework.

An investigation into how to assess the quality of education through curriculum intent, implementation and impact
December 2018, No. 180035

3

- regular curriculum review
- using the curriculum as the progression model
- intelligent use of assessment to inform curriculum design
- revisiting and recalling previously learned knowledge 'baked into' the curriculum planning
- distributed curriculum leadership.

However, these were only assumptions based on leaders' perceptions of what matters when it comes to their curriculum design. This suggested that further research was required to establish more precisely whether these factors are closely related to curriculum quality or not.

Our intention for phase 3, therefore, was to design a research model in which the curriculum intent discussion, corroborated by first-hand evidence of curriculum implementation, could inform a series of indicators on the extent of curriculum quality. This model would then be trialled in a range of schools by a small group of Her Majesty's Inspectors (HMI) to test for validity.

The focus of the research model was on building an evidence base that allows us to refine the most useful aspects for future use on inspection. The model itself is not intended for direct use on inspection. The findings from this investigation has instead provided a research basis for and will feed into the design of the EIF. A summary of the main findings from the phase 3 study can be found in Her Majesty's Chief Inspector's commentary that complements this report.[3]

Method

General overview

Having established a viable process for collecting evidence on school leaders' curriculum intent from the phase 2 study, we sought to understand how this fits into a broader model of assessing not only intent, but also curriculum implementation and impact. The specific research question we were looking to answer was 'How to assess intent, implementation and impact in a research model of curriculum quality?' From this, we generated a series of objectives for the phase 3 investigation:

- to develop a series of curriculum indicators in a research model and establish whether they have a valid relationship with curriculum quality

- to test specific inspection practices to determine a valid method for capturing and understanding curriculum quality

---

[3] HMCI's commentary: Commentary on curriculum research - phase 3. Ofsted, December 2018; www.gov.uk/government/speeches/commentary-on-curriculum-research-phase-3.

An investigation into how to assess the quality of education through curriculum intent, implementation and impact
**December 2018, No. 180035**

4

- to understand whether the indicators and inspection practices allow inspectors to distinguish between curriculum intent and implementation
- to determine the practical limitations of what might be possible in the context of a routine inspection.

The research design for phase 3 focused on two core principles:

- developing a valid research model of the quality of education, covering the intention, implementation and impact of the curriculum
- designing an inspection approach that focused on the collection of first-hand evidence on the quality of education.

We expected that alignment between these principles – the method of evidence collection and the systematic structure of the research model – would enhance the validity of the model.

Figure 1: Design approach to the phase 3 study
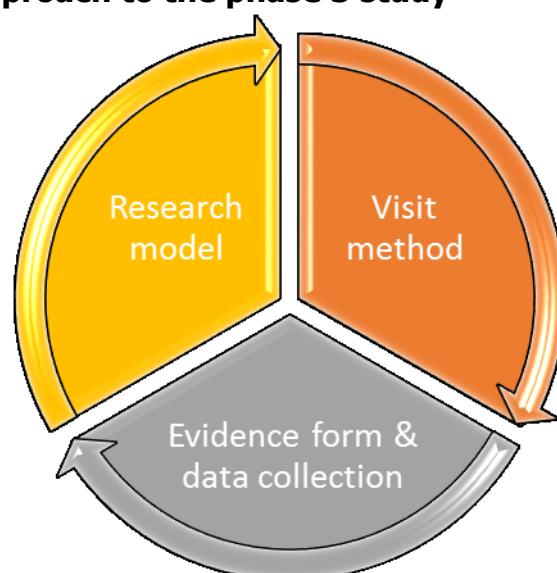
Nine HMI were available for the fieldwork of this study and covered all eight Ofsted regions. Two of these inspectors were designated HMI leads for the whole project and worked closely with the senior research lead. They contributed to the development of the research design, methodology and inspector training, as well as participating in the fieldwork. On each visit, inspectors were expected to carry out the following:

- examine each school's unique curriculum offer
- use the school's own curriculum to examine the extent and success of curriculum implementation in partnership and alongside middle and senior leaders
- consider the impact of 'leaders' deliberate action' to implement their curriculum, particularly in the last 18 months

An investigation into how to assess the quality of education through curriculum intent, implementation and impact
December 2018, No. 180035

5

- conduct a series of activities alongside school staff to look at first-hand evidence of curriculum implementation

- examine the progression in learning that the pupils experience at the school by asking leaders to share the school's curriculum and what pupils learn from Year 1 to Year 6

- use banding criteria to allot a rating of best fit to the quality of curriculum.

The following sections discuss the content of the research model and the visit process in more detail.

Research model

The phase 2 research looked at a small number of schools that we had reason to believe were focused more than most schools on their curriculum design and management. The findings from this work provided insight into the dimensions of curriculum that were routinely specified by leaders as being part of the curriculum design across these schools. This included:

- local context and filling the gaps from pupils' backgrounds

- focus on subject disciplines even when topics are taught

- considering depth and breadth of curriculum content

- seeing the curriculum as the progression model

- having a clear purpose for assessment

- reviewing and evaluating curriculum design

- clear curriculum leadership (often distributed) and ownership.

On the basis of these factors, we developed a set of curriculum indicators to test whether it was possible to accurately assess curriculum quality across a range of different school types. Considering the purposive nature of the phase 2 sample, it was especially important to trial any such curriculum indicators in a more diverse set of schools to see whether they work in different contexts. In this way, we hypothesised that the indicators would:

- work in different school contexts without biasing against a particular type of school or pupil cohort

- distinguish between effective and ineffective curriculum design

- pick up where curriculum narrowing is happening, as identified in the phase 1 study

- identify differences between curriculum intent as specified by leaders and curriculum implementation.

Using the factors from the phase 2 research as a starting point, we developed 25 indicators that we expected to be associated with curriculum quality. The development of these indicators was informed further by the available research

An investigation into how to assess the quality of education through curriculum intent, implementation and impact
December 2018, No. 180035

6

literature and the experience of several HMI. A limitation of the study is that indicator development focused on generic aspects of curriculum quality and not subject-specific aspects. In time, we would like to be able to use subject-specific curriculum indicators, because we think that this will be even more powerful in assessing the quality of education. The full list of indicators can be viewed in figure 2. Indicators in orange are indicators framed around curriculum intent; those in green are implementation indicators; and those in blue relate to impact.

An investigation into how to assess the quality of education through curriculum intent, implementation and impact
December 2018, No. 180035

7

Figure 2: List of curriculum indicators in the research model

No.	Indicator
1a	There is a clear and coherent rationale for the curriculum design
1b	Rationale and aims of the curriculum design are shared across the school and fully understood by all
1c	Curriculum leaders show understanding of important concepts related to curriculum design, such as knowledge progression and sequencing of concepts
1d	Curriculum coverage allows all pupils to access the content and make progress through the curriculum
2a	The curriculum is at least as ambitious as the standards set by the national curriculum/external qualifications
2b	Curriculum principles include the requirements of centrally prescribed aims
2c	Reading is prioritised to allow pupils to access the full curriculum offer
2d	Mathematical fluency and confidence in numeracy are regarded as preconditions of success across the national curriculum
3a	Subject leaders at all levels have clear roles and responsibilities to carry out their role in curriculum design and delivery
3b	Subject leaders have the knowledge, expertise and practical skill to design and implement a curriculum
3c	Leaders at all levels, including governors, regularly review and quality assure the subject to ensure that it is implemented sufficiently well
4a	Leaders ensure that ongoing professional development/training is available for staff to ensure that curriculum requirements can be met
4b	Leaders enable curriculum expertise to develop across the school
5a	Curriculum resources selected, including textbooks, serve the school's curricular intentions and the course of study and enable effective curriculum implementation
5b	The way the curriculum is planned meets pupils' learning needs
5c	Curriculum delivery is equitable for all groups and appropriate
5d	Leaders ensure that interventions are appropriately delivered to enhance pupils' capacity to access the full curriculum
6a	The curriculum has sufficient depth and coverage of knowledge in the subjects
6b	There is a model of curriculum progression for every subject
6c	Curriculum mapping ensures sufficient coverage across the subject over time
7a	Assessment is designed thoughtfully to shape future learning. Assessment is not excessive or onerous
7b	Assessments are reliable. Teachers ensure systems to check reliability of assessments in subjects are fully understood by staff
7c	There is no mismatch between the planned and the delivered curriculum
8	The curriculum is successfully implemented to ensure pupils' progression in knowledge – pupils successfully 'learn the curriculum'
9	The curriculum provides parity for all groups of pupils

We also developed detailed guidance on how inspectors should assess the indicators. This took the form of a detailed rubric to guide inspectors in making informed assessments against each indicator on a 1–5 scale (with 5 equating to a strong curriculum design and 1 a weak design). The rubric provided a systematic structure around the indicator design so that inspectors could make consistent evaluations in determining the strengths and weaknesses of a school's curriculum.

Evidence relating to a score of 5 on a curriculum indicator would show exceptional curriculum features, whereas a score of 2 would indicate that there are major weaknesses in this aspect of curriculum design and leaders' work to remedy this is either very recent or having little impact. Evidence allotted to a score of 1 would demonstrate that this aspect of curriculum is absent. Figure 3 provides further details of the categories each number represents on the five-point scale.

Figure 3: Categories applied in the rubric for scoring the curriculum indicators reliably

5	4	3	2	1
This aspect of curriculum underpins/is central to the school's work/embedded practice/may include examples of exceptional curriculum	This aspect of curriculum is embedded with minor points for development (leaders are taking action to remedy minor shortfalls)	Coverage is sufficient but there are some weaknesses overall in a number of examples (identified by leaders but not yet remedying)	Major weaknesses evident in terms of either leadership, coverage or progression (leaders have not identified or started to remedy weaknesses)	This aspect is absent in curriculum design

We adopted the 1–5 scale as a means to reduce inspector unconscious bias. For instance, a four-point scale was associated too closely with the routine scale applied on inspection. By also reversing the order, inspectors needed to think more deeply about the evidence they collected and how this fits into the rubric, rather than directly attributing this in terms of an inadequate or outstanding curriculum. The five-point scale also provided a more subtle way of measuring the variability in curriculum design across schools, which enhanced our post-visit analysis of the data collected.

The indicators and rubric were designed to be deliberately lengthy and contain some duplication. We included more indicators than necessary – or possible for us to look at – with the intention that post-visit data analysis would allow us to narrow and refine the indicators to just those that are clearly related to curriculum quality and can be used reliably. We also expected some indicators to naturally fall out of the research model through inspectors finding it difficult to apply and collect data on all the indicators in practice. We collected inspector feedback after the completion of the fieldwork to corroborate their views with the data analysis to assist with the refinement of the indicators.

Visit process

The evidence collected for the research model is based on both qualitative retrieval and quantitative assessment. Inspectors needed to apply their professional judgement to make informed assessments on the evidence collected and to determine how this met the criteria of the rubric and the five-point scale. Therefore, alongside the indicators and rubric, we also designed a systematic process for collecting the evidence. This was developed to provide some assurances that inspectors could gather information from the research visits in a uniform way. This would also enhance inspector reliability in how they assess the curriculum.

In general, the visit design included a series of meetings taking place across the school day, although these sessions often involved more than a straightforward discussion. For instance, a large part of the subject leaders meeting involved the inspector carrying out a work scrutiny with them, including looking at pupils' workbooks. The collection of first-hand evidence to understand the effectiveness of curriculum implementation was crucial here, especially to triangulate with the expressed curriculum intent of school leaders. To cover enough of the curriculum, inspectors scrutinised the quality of at least four subject areas per school. One subject was English or mathematics (not both) because, particularly in primary schools, this was expected to act as the baseline to assess against the three foundation subjects they looked at.

As part of the design, inspectors engaged with four specific groups:

- senior leaders (intent)
- subject leads/teachers (implementation)
- pupils (implementation and impact)
- governors (intent and impact)

Inspectors were given a set of standardised questions and prompts to apply when carrying out the activities with each group. The standardised approach was also reflected in the evidence forms created for data collection. These questions ensured that detailed coverage could be collected for all 25 curriculum indicators. The questions were also adapted for specific groups of staff and for pupils. Further details on the methodology applied for each group can be found in Annex A.

Overall banding criteria

Owing to the focus on four subject areas for each visit, we also asked inspectors to provide an overall banding score on a five-point scale to reflect the overall curriculum quality of the school. This allowed inspectors to assess overall quality of education when they noticed variability between the subjects reviewed, as well as where differentiation existed between intent and implementation. The overall banding criteria was from the main rubric and required inspectors' professional judgement to apply the evidence accurately to the criteria. The inclusion of this metric was also useful as a dependent variable for the post-visit analysis of the quantitative data collected.

Sampling

As discussed, the phase 3 research design builds on the findings from the phase 2 research. Part of the reason for doing the phase 3 investigation was extrapolating the curriculum design factors identified from phase 2 to see whether they are equally applicable across a broad range of schools and not just those known to be involved in curriculum design. Therefore, the sample for phase 3 was designed to be balanced in order to test the validity of our curriculum research model across differing school contexts.

We selected an initial sample of schools using the following criteria:

- previous inspection judgements (outstanding, good and RI only)
- geographical location (Ofsted regions)
- school type (LA maintained/academies)

The sample was selected to closely match the national level, where appropriate data existed. However, this does have some implications at the regional level, because equal distribution of the criteria across Ofsted regions was not factored into the sample design. We also over-sampled for secondary schools and schools that were judged outstanding or requires improvement at their last routine inspection to ensure that a greater variety of schools would be included in the sample.

The sample was checked against other criteria to minimise any unintended selection bias. This ensured that we would be able to visit a similar proportion of affluent and deprived schools across the sample and that we included a balance of schools with strong or weak performance data. We were also keen to see whether the curriculum indicators were applicable to very small primary schools, so cohort size was also included in balancing the sample to a degree.

Some schools in the original sample selection were replaced because of leaders not wishing to participate in the research or being unavailable to participate on dates when inspectors were available for the fieldwork. It was not always straightforward to find a statistically similar school as a replacement. In these cases, we used a close fit as an alternative instead to maintain sample balance. In total, leaders from eight schools were unable to arrange a convenient date for the research visit or decided not to participate.

Overall, 64 schools were selected for the research visits in the summer term 2018. This consisted of 33 primary schools, 29 secondary schools and two special schools. Typically, one inspector was on site for each visit. While the 64 schools may not necessarily provide full coverage across all school types, we still expected to test whether the indicators and rubric work equally well across differing contexts. The two special schools were included to further test the validity of the indicators on this basis. Details of the sample balance achieved for the 64 schools involved in the fieldwork can be found in Annex B.

The subjects selected focused on compulsory national curriculum subjects. The study attempted to cover a broad balance of subjects across the sample beyond English and mathematics. However, this did not always work. Staff absence or leader preferences on the subjects they wished inspectors to look at led to a slight imbalance in the subjects assessed across the research visits.

Pilots

Four pilot visits were carried out by the senior research lead and the two designated HMI leads on the project prior to the full fieldwork being completed by other inspector colleagues. The pilots were conducted to ensure the integrity between the indicators, evidence gathering process and the evidence form. This resulted in some tweaking of the indicators, rubric and the visit process, but generally the pilot visits highlighted that the design was viable and could be implemented in practice. The results from all four pilots are, therefore, included in the full sample of 64 schools.

Inspector training

It was essential that inspectors were trained on how to apply the research model (indicators and rubric) and how it aligns with the visit process so that they could carry out the fieldwork in a reasonably consistent way. An intensive one-day training session clarified the model design and visit process so that inspectors could confidently apply this during the fieldwork. Each inspector's first visit was also attended by one of the designated HMI leads on the project. This was so that the HMI leads, who had developed the process and trialled it in the pilots, could model the process to their colleagues for their first couple of subject meetings.

The HMI leads also quality-assured the evidence forms from all the visits carried out during the study. This was to provide assurance that the evidence being collected was in line with the research methodology, particularly the content of the rubric and the question prompts devised. Generally, quality assurance of the earliest evidence bases from the fieldwork and the feedback provided meant that we could make improvements in the quality of data collected during the later research visits.

Limitations

While the methodology for the study is quite extensive, there remain a few limitations to its design that have implications for external validity.

- The indicators developed for the study are focused on generic attributes of curriculum quality, not subject-specific content. Given our inspector resource, we are not yet at the stage to test out subject-specific aspects of curriculum quality.

- Lesson observation was not an in-depth or central part of the triangulation process – the work scrutiny and pupil discussions carried more weight. This was beneficial in ensuring that inspectors looked at the quality of learning over time and not individual lessons. This is not intended to devalue the purpose of observation and we expect it to carry considerable importance for helping to assess the quality of education in the new framework, in

terms of the contribution that lessons make to realising a school's intent for its curriculum. As such, we will be trialling a lesson observation model in a separate study that complements this one.

■ There is an argument as to whether the focus on four subjects is sufficient for making valid assessments on curriculum quality. We could have looked at more subjects per visit if each school visit was spread over two days. However, based on the resource available for the project, this would have meant constricting our overall sample size as a consequence. As the main purpose of the study was to test the research model in a varied range of school provision, we decided that the four subject design better met this objective.[4] Furthermore, looking at four subjects did allow us to go into some depth and gain genuine understanding of those subjects, which seemed more valuable than trying to gain a surface-level picture of a greater number of subjects.

■ There is potentially a level of unreliability that we cannot control for in the study, because we were not able to test for interrater reliability between inspectors in the study design. However, the systematic design of the research model, inspector training and the quality assurance processes in place should limit this.

Evaluation

Following the completion of the fieldwork, we collated the scores from the evidence forms of the schools visited into a data-set. We then used this to carry out a number of analyses to test whether the indicators are sufficiently valid. We aimed to do this in two ways. First, we established the face validity of our research model in consideration of our initial hypotheses and the research evidence that already exists. As a large body of literature explains, curriculum narrowing is a recent phenomenon in Western education systems.[5] We assumed that our research model would also be able to identify such concerns.

Second, we looked at the statistical relationships between the indicators. We hypothesised that the most valid indicators in our model of curriculum quality could be identified. This is an important task given the number of indicators included in

---

[4] In one school we did get the opportunity to review seven subject areas as part of the research visit. The two inspectors involved on this visit indicated that this told us no more than what we were getting from four subject areas.

[5] For instance:

■ Berliner, D (2011). Rational responses to high stakes testing: the case of curriculum narrowing and the harm that follows. Cambridge Journal of Education, 41(3)

■ Polesel, J., Rice, S., & Dulfer, N. (2014). The impact of high-stakes testing on curriculum and pedagogy: a teacher perspective from Australia. Journal of Education Policy, 29(5)

■ Barnes, J., & Scoffham, S. (2017). The humanities in English primary schools: struggling to survive. Education 3-13, 45(3)

■ Ofsted (2017), HMCI's commentary: recent primary and secondary curriculum research, https://www.gov.uk/government/speeches/hmcis-commentary-october-2017

the research model. For routine inspection, we would not expect to provide inspectors with such a large number of indicators, or the extensive rubric to go with them, given the short timescales of modern inspection practice. Refining the indicators to those that matter most was therefore paramount.

What does the overall curriculum quality rating tell us about the validity of the research model?

Figure 4 shows that the primary schools visited generally had weaker aspects of curriculum quality. We assessed a larger proportion of secondary schools to have a curriculum quality of band 4 or 5, with only three schools in band 1 or 2. Inspectors identified that almost half of the primary schools had a curriculum quality in the lowest two bands, although there were also two primary schools that scored in the highest band.

Figure 4: Curriculum overall banding by school phase

School phase	Band 1	Band 2	Band 3	Band 4	Band 5	Total
Primary	3	12	10	6	2	33
Secondary	-	3	10	17	1	31
Total	3	15	20	23	3	64

Band 1 = curriculum design is absent, band 5 = curriculum design is central to the school's work and may include examples of exceptional curriculum.
The two special schools are included in the secondary school data.

The subject level data collected suggests variation in subject quality may be partially responsible for contributing to this pattern. Unsurprisingly, owing to the prominence that English and mathematics have in the performance tables, the first-hand evidence collected from the primary schools shows a large extent of variability between successful curriculum implementation of English and mathematics compared with foundation subject areas (figure 5).

In most of the foundation subject departments, a greater proportion were found to be meeting the criteria for band 1 and 2 of the rubric. This includes science, a core subject, which adds further evidence to the perception that this subject has become downgraded compared with English and mathematics since the scrapping of the key stage 2 test.

Figure 5: Indicator 6a by subject departments assessed during the 33 primary schools visits

Subject area	Band 1	Band 2	Band 3	Band 4	Band 5	Total
Core						
English	-	1	6	9	1	17
Maths	1	1	8	6	1	17
Science	2	4	6	1	1	14
Foundation						
Humanities	7	7	11	5	-	30
Arts	4	9	6	2	1	22
PE	1	-	2	6	1	10
Technology	6	4	4	2	2	18
MFL	-	-	2	1	3	6
Total	21	26	45	32	10	134

Band 1 = curriculum design is absent, band 5 = curriculum design is central to the school's work and may include examples of exceptional curriculum.
Indicator 6a: The curriculum has sufficient depth and coverage of knowledge in subjects
Includes two subject reviews conducted in the primary phase of an all through school.
Technology includes computer studies.

Evidence from the visits showed that in the weaker primary schools, the main focus was on putting a plan together, but not checking the implementation of that plan effectively enough. This was linked to an onus on delivering the content of the national curriculum for foundation subjects, but without careful thought given to the progression of knowledge and skill that would make this useful learning for pupils. In some instances, there was more of a tick-box approach for ensuring that specific content had been delivered on the curriculum plan to meet national curriculum objectives, rather than ensuring that pupils had understood and could apply this content to future learning. So, while content was being delivered across the foundation subjects, this was often poorly organised and sequenced and lacking sufficient oversight from senior leaders.

Timetabling was also a concern, particularly the blocking of subjects over a specific timeframe. This could mean pupils only doing science, history or geography one term per year. In many of the schools visited, this made it very difficult for pupils to progress in the subject. This practice often led to a lack of coherence as the work between year groups on some foundation subjects consisted of piecemeal content where important concepts were underdeveloped. Evidence from the visits indicate that staff subject knowledge was of considerable importance, particularly when it came to designing appropriate progression through content.

Although the data appears positive for modern foreign languages in primary schools, inspectors had fewer discussions with subject leads than for other subjects in the

study. In some cases, this was due to staff absence meaning that it was not possible to focus on languages. Therefore, conclusions are more difficult to draw for this subject.

In comparison, figure 6 shows that the subject level data for secondary schools was more impressive.

Figure 6: Indicator 6a by subject departments assessed during the 29 secondary schools visits

Subject area	Band 1	Band 2	Band 3	Band 4	Band 5	Total
Core						
English	1	-	3	6	6	16
Maths	-	-	3	6	5	14
Science	-	-	2	4	2	8
Foundation						
Humanities	4	2	6	6	8	26
Arts	-	1	2	4	6	13
PE	-	1	1	1	3	6
Technology	-	2	2	2	2	8
MFL	1	-	3	4	-	8
Other	2	-	5	4	4	15
Total	8	6	27	37	36	114

Band 1 = curriculum design is absent, band 5 = curriculum design is central to the school's work and may include examples of exceptional curriculum.
Indicator 6a: The curriculum has sufficient depth and coverage of knowledge in subjects
One of the secondary schools visited was an all through school. Only two subject reviews were carried out in the secondary phase for this visit.
Technology includes computer studies.

Compared to the primary schools visited, there was much less difference between the quality of implementation of the foundation subjects and the core subjects. Nearly all the secondary schools demonstrated considerable strengths in the implementation of the curriculum in three or four of the subject disciplines examined. Only a few demonstrated at least two subject areas with major weakness in terms of either subject leadership, coverage or progression.

However, curriculums in some subjects were still being implemented weakly compared with those for English and mathematics. For instance, in a few of the modern foreign languages departments, many of the key features of successful curriculum design, planned progression and implementation were limited due to the lack of availability of subject specialists. History was also neither organised nor implemented well in a number of schools, often to the detriment of a clear

progression through the curriculum. Again, the lack of subject expertise, especially in leadership roles, contributed to these weaknesses.

The curriculum quality data appears to have implications for the outstanding exemption. While a correlation exists between curriculum quality and the latest overall effectiveness judgement for the schools visited (figure 7), inspectors still identified concerns about the curriculum in outstanding and good schools. The band 2 and 3 definitions in the rubric – on the assumption that our research model is valid – suggest that nine outstanding schools did not have a high-quality curriculum offer in place for their pupils at the time of the research visits. Of these schools, three were last inspected over 10 years ago and two had their last inspection in 2012/13. Again, this pattern was particularly being driven by the primary schools in the sample (figure 8).

Figure 7: Curriculum overall banding by the overall effectiveness judgement of the schools visited at their last routine inspection

Overall effectiveness	Band 1	Band 2	Band 3	Band 4	Band 5	Total
Outstanding	-	2	7	7	3	18
Good	2	7	9	12	-	30
Requires improvement	1	6	4	4	-	15
Total	3	15	20	23	3	64

Band 1 = curriculum design is absent, band 5 = curriculum design is central to the school's work and may include examples of exceptional curriculum.
Overall effectiveness judgements based on data at time of sampling

Figure 8: Curriculum overall banding by the overall effectiveness judgement of the primary and secondary schools visited

Overall effectiveness	Band 1	Band 2	Band 3	Band 4	Band 5	Total
Primary						
Outstanding	-	2	4	1	2	9
Good	2	6	5	4	-	17
Requires improvement	1	4	1	1	-	7
Secondary						
Outstanding	-	-	3	6	1	10
Good	-	1	4	8	-	13
Requires improvement	-	2	3	3	-	8
Total	3	15	20	23	3	64

Band 1 = curriculum design is absent, band 5 = curriculum design is central to the school's work and may include examples of exceptional curriculum.
Overall effectiveness judgements based on data at time of sampling
The two special schools are included in the secondary school data.

One area we were interested in about the research model was its relationship with performance data. In particular, we wanted to know whether the indicators could tell us something about a school's curriculum that is independent of a school's published attainment and progress data. Figures 9 and 10 show that the model is able to unpick this.

Figure 9: Curriculum overall banding by key stage 2 and key stage 4 progress measure bandings

Progress	Band 1 or 2	Band 3	Band 4 or 5	Total
Above average	3	4	8	15
Average	5	10	12	27
Below average	3	3	2	8
No data	7	3	2	12
Total	18	20	24	62

Band 1 = curriculum design is absent, band 5 = curriculum design is central to the school's work and may include examples of exceptional curriculum.
Key stage 4 progress 8 data (2018) and key stage 2 mathematics data (2017) have been merged for statistical disclosure control purposes.
The progress bandings shown are based on the five progress measure bandings calculated by the Department for Education. 'Above average' combines 'above average' and 'well above average'. 'Below average' combines 'below average' and 'well below average'.
Proportion of schools in each banding differs between the two progress measure. The key stage 4 progress measure encompasses more subjects than the single-subject measures used at key stage 2.
Data for the two special schools is not included.
Schools with no data are newly opened schools or infant schools.

Figure 10: Curriculum overall banding by KS2 and KS4 attainment

Pupil attainment 2017/18	Band 1	Band 2	Band 3	Band 4	Band 5	Total
Above or in line with average	1	5	8	14	2	30
Below average	2	4	8	9	1	24
No attainment data	-	6	4	-	-	10
Total	3	15	20	23	3	64

Band 1 = curriculum design is absent, band 5 = curriculum design is central to the school's work and may include examples of exceptional curriculum.
The key stage 2 attainment measure used is the percentage of pupils reaching expected standard in reading, writing and maths. The key stage 4 attainment measure used is the attainment 8 score.
Schools with no data are newly opened schools or infant schools.

As we might expect, while a relationship exists between curriculum quality and pupil progress and attainment scores, the data from the school visits identifies some variations. For instance, some schools with high progress or attainment scores were

assessed by inspectors as having a weak curriculum offer. Even more importantly, some schools with below-average data were deemed to have a strong curriculum design in place. The curriculum quality scores given by inspectors, therefore, reflected the criteria of the rubric more so than the published performance data. This suggests that by looking at the curriculum we can move beyond performance data and identify evidence on what is being taught to complement the performance data.

We also wanted to know whether the research model was equally applicable across variable school contexts. The bandings shown in figure 11 suggest that this is possible when the level of deprivation is considered. The data shows little difference in curriculum quality between those schools visited in the most and least deprived IDACI quintiles, for instance. Curriculum quality, therefore, does not appear to be correlated with deprivation.

Figure 11: Curriculum overall banding by the IDACI quintile of each school

IDACI	Band 1	Band 2	Band 3	Band 4	Band 5	Total
Quintile 1 (least deprived)	1	4	4	4	-	13
Quintile 2	-	4	5	3	2	14
Quintile 3	-	-	5	5	1	11
Quintile 4	1	4	3	5	-	13
Quintile 5 (most deprived)	1	3	3	6	-	13
Total	3	15	20	23	3	64

Band 1 = curriculum design is absent, band 5 = curriculum design is central to the school's work and may include examples of exceptional curriculum.
Deprivation is based on the Income Deprivation Affecting Children Index (IDACI). The deprivation of a provider is based on the mean of the deprivation indices associated with the home post codes of the pupils attending the school rather than the location of the school itself. The schools are divided into five equal groups (quintiles), from 'most deprived' (quintile 5) to 'least deprived' (quintile 1).

This is encouraging because it suggests that having a deprived intake is not a barrier to offering a rich and broad curriculum to pupils, even if this is not reflected as clearly in attainment and progress data. Conversely, it also suggests that some schools in more affluent areas are providing a low-quality curriculum offer to their pupils or gaming or coasting on the back of more affluent pupil intakes. Furthermore, this suggests that a move away from using performance data as a large part of the basis for judgement and towards focusing on curriculum quality will allow us to reward schools in challenging circumstances much more equitably.

Finally, we looked at whether being an academy or free school made a difference to the quality of curriculum on offer. Figure 12 shows, on the basis of our sample, that there was little variation in the curriculum quality between academies and maintained schools.

Figure 12: Curriculum overall banding by school type

School type	Band 1	Band 2	Band 3	Band 4	Band 5	Total
Academy or free school	-	8	9	12	2	31
LA maintained	3	7	11	11	1	33
Total	3	15	20	23	3	64

Band 1 = curriculum design is absent, band 5 = curriculum design is central to the school's work and may include examples of exceptional curriculum.

Overall, these descriptive statistics point towards our research model having good face validity. In terms of our stated assumptions, the model works in different contexts and the data indicates it is not biased towards any particular type of context. It is also able to distinguish between effective and ineffective curricular design. In particular, the variation in the data collected at the subject level suggests that implementation is dependent on a range of factors, with the rubric and the related inspection practice able to unpick this. The model is also able to identify weaknesses in curriculum design despite other data providing an alternative view on pupil outcomes. Finally, the model is also effective in identifying where curriculum narrowing is happening, particularly in primary schools. Importantly, the empirical evidence supports the findings from phase 1 of our study and indicates that inspectors can establish this more easily through the method of scrutiny applied on these visits.

The next section will look at the last of the assumptions we made on how well the model (and, by extension, inspectors) can identify differences between curriculum intent and curriculum implementation.

Can inspectors distinguish between curriculum intent and curriculum implementation using the research model?

We carried out a number of statistical analyses of our research model. These analyses showed that there are two main factors, one relating to intent and one to implementation, and that a number of predictors are strongly associated with each of the factors. These analyses were also corroborated by the views of the HMI who carried out the fieldwork to consolidate the main predictors. Figure 13 visualises this model of curriculum quality and the relationship between intent and implementation:

Figure 13: Curriculum quality model, based on evidence from statistical analysis and HMI feedback
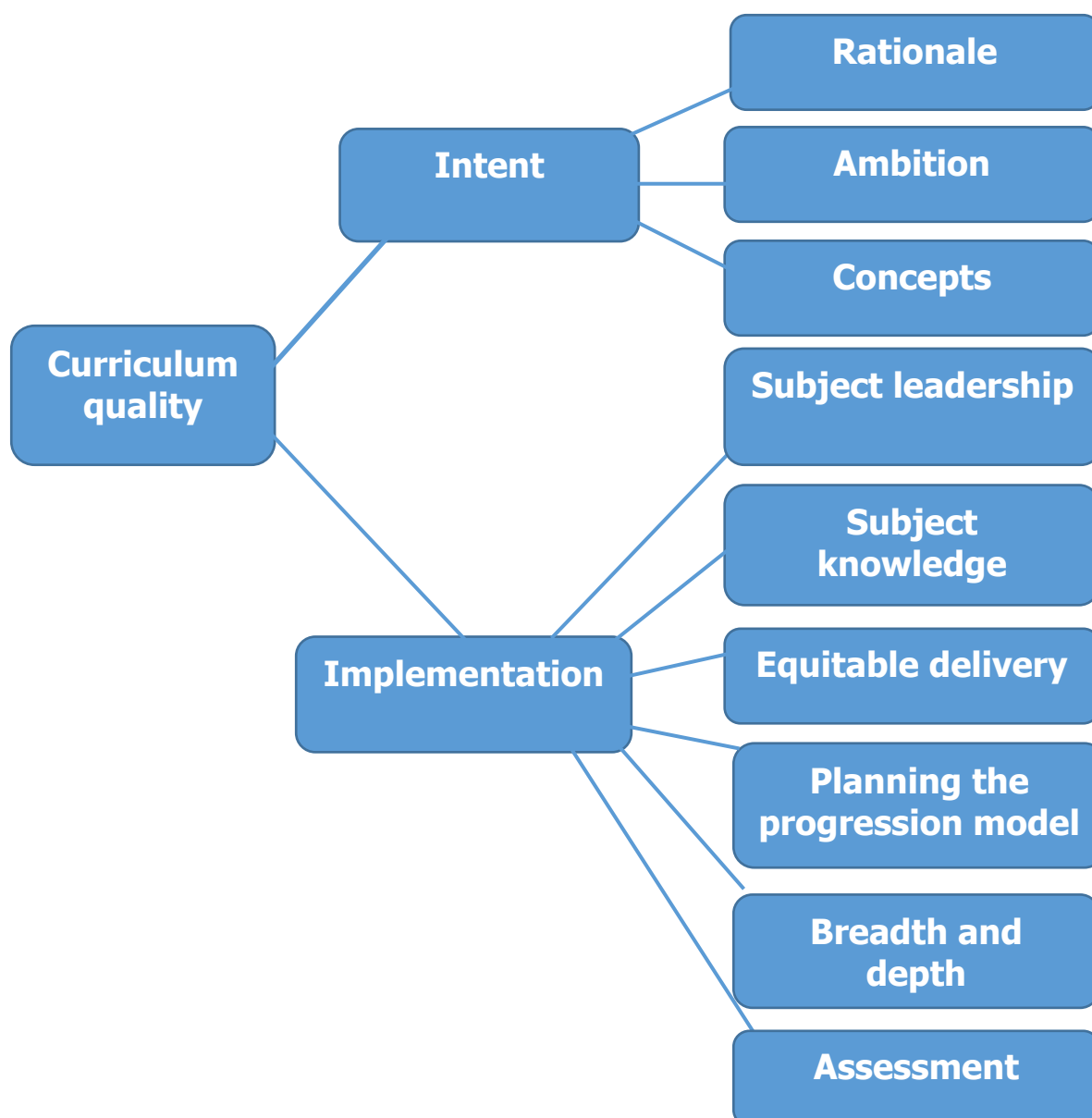
It is important to note what we mean by 'concepts' and 'planning'. The evidence from the visits highlight that 'concepts' relate to senior and middle leaders having an understanding of curriculum design. For instance, they have a secure understanding that the curriculum should be logically organised and sequenced to aid pupils' learning and, crucially, can exemplify this. It was less likely that effective curriculum implementation was embedded without leaders first articulating an understanding of this concept.

The evidence also indicates that effective curriculum 'planning' was not just about having a written plan of the content being taught. In the bands 4 and 5, schools' planning focused more on, for instance, how curriculum leaders ensured that:

- content is sequenced to ensure that components of knowledge lead to conceptual understanding

- opportunities for pupils to practise what they knew – so they could deepen their understanding in a discipline – were built into the curriculum

- the layering of knowledge and concepts were secure so that pupils could make progress in the curriculum form their starting points.

The data supporting this model of curriculum quality is set out below.

Figures 14 and 15 show how the ratings for the intent and implementation indicators were distributed across the school visits. This demonstrates that the majority of schools achieved a score of band 3 or higher. A smaller proportion of schools were given the lowest two scores. This means that the indicators are distributed in a similar pattern to that of the current overall effectiveness grade.[6]

However, the interesting aspect of these two charts is the effect that the averaging of the individual subject scores has on the implementation indicators (3a to 7c). Compared with the intention indicators (1a to 2d), there is a slight drop in the proportion of schools scoring particularly high or low on the implementation indicators, with a corresponding increase in the frequency with which inspectors gave band 3. This shows that for some schools inspectors were changing their views on the quality of curriculum from the senior leaders' discussion once the first-hand evidence from the practical implementation of the curriculum had been considered.

---

[6] State-funded schools inspections and outcomes as at 31 August 2018, Ofsted; 2018
www.gov.uk/government/statistics/state-funded-schools-inspections-and-outcomes-as-at-31-august-2018

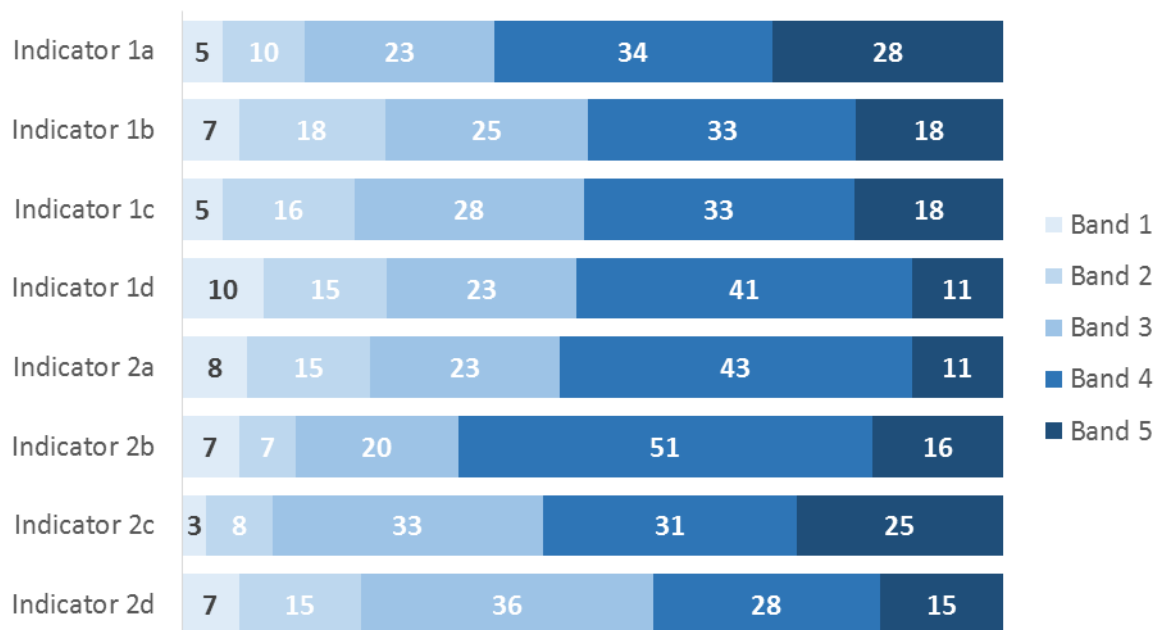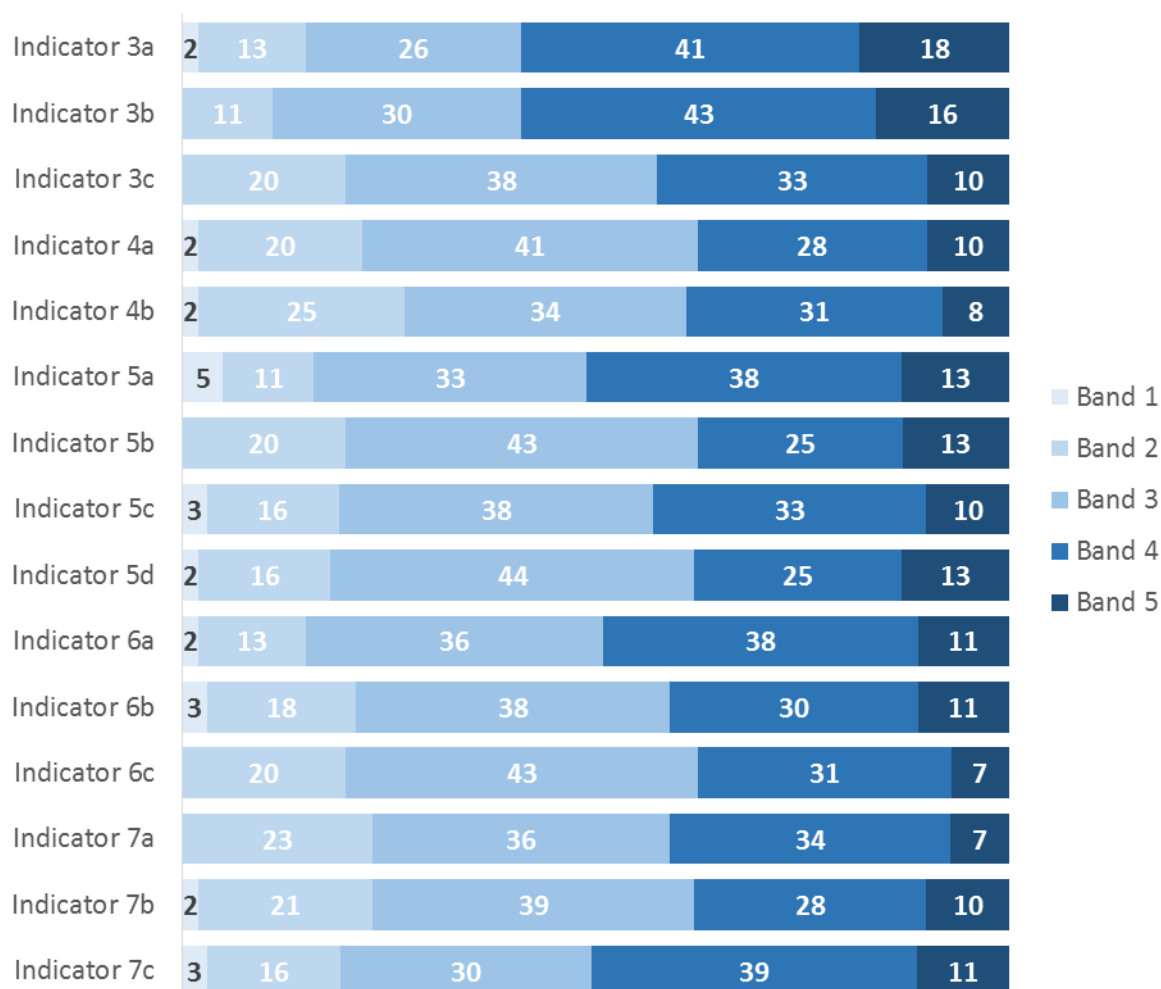## Figure 14: Percentage of schools achieving each intent indicator rating

Indicator	Band 1	Band 2	Band 3	Band 4	Band 5
Indicator 1a	5	10	23	34	28
Indicator 1b	7	18	25	33	18
Indicator 1c	5	16	28	33	18
Indicator 1d	10	15	23	41	11
Indicator 2a	8	15	23	43	11
Indicator 2b	7	7	20	51	16
Indicator 2c	3	8	33	31	25
Indicator 2d	7	15	36	28	15

Figure 15: Percentage of schools achieving each implementation indicator rating

Indicator	Band 1	Band 2	Band 3	Band 4	Band 5
Indicator 3a	2	13	26	41	18
Indicator 3b		11	30	43	16
Indicator 3c		20	38	33	10
Indicator 4a	2	20	41	28	10
Indicator 4b	2	25	34	31	8
Indicator 5a	5	11	33	38	13
Indicator 5b		20	43	25	13
Indicator 5c	3	16	38	33	10
Indicator 5d	2	16	44	25	13
Indicator 6a	2	13	36	38	11
Indicator 6b	3	18	38	30	11
Indicator 6c		20	43	31	7
Indicator 7a		23	36	34	7
Indicator 7b	2	21	39	28	10
Indicator 7c	3	16	30	39	11

Figures are rounded and may not add to 100.

In order to get underneath this observation of the data further, we investigated whether there were any underlying factors of the 25 indicators. This analysis identified two main factors. Figure 16 shows the factor loadings, which produced one factor that aligns very closely with the indicators grouped under the 'implementation' category and a second factor that aligns very closely with the indicators grouped under the 'intent' category.[7] This suggests that the initial groupings of the indicators within the research model have been applied correctly and that they are investigating what they were initially set out to do.

---

[7] The initial scree plot suggested that there were two factors underlying the data. Analysis was run to explore these further, which identified there were several high loadings onto each of the two factors, reflecting the high correlations between our indicators. To reduce the overall numbers of indicators, a high loading cut off of 0.70 was applied.

Figure 16: Results of factor analysis conducted on the 25 quality indicators

|	Factor 1	Factor 2
Indicator 1a		0.92
Indicator 1b		0.91
Indicator 1c		0.78
Indicator 1d		0.80
Indicator 2a		0.70
Indicator 2b		
Indicator 2c		
Indicator 2d		0.80
Indicator 3a		
Indicator 3b		
Indicator 3c	0.72	
Indicator 4a	0.71	
Indicator 4b	0.71	
Indicator 5a		
Indicator 5b	0.88	
Indicator 5c		
Indicator 5d	0.81	
Indicator 6a	0.99	
Indicator 6b	0.97	
Indicator 6c	0.96	
Indicator 7a	0.99	
Indicator 7b	0.92	
Indicator 7c	0.86	
Indicator 8		
Indicator 9		

Indicators 1a to 2d = intent, 3a to 7c = implementation, 8 and 9 = impact.

Further analysis of just the intent and implementation indicators, but not the two impact ones, also support this conclusion. In this case, the factor analysis automatically grouped the indicators into their respective intent and implementation categories as well.

In addition, we generated a factor plot to visualise how each individual school scored on the two different factors. This was coded to show each school by their phase, because the descriptive statistics suggest that this variable had some influence on

the overall curriculum scores. The pattern in figure 17 again identifies the phase influence. That is, primary schools in our sample typically featured weaker curriculum design, especially around implementation (see the bottom left quadrant). The top right quadrant shows schools with both strong curriculum intent and implementation, which again was more common in our secondary school sample. However, the chart also reflects on inspectors' ability to apply the intent and implementation structure of the methodology appropriately.

Figure 17: Factor loadings for the intent and implementation indicators at the individual school-level, grouped by phase
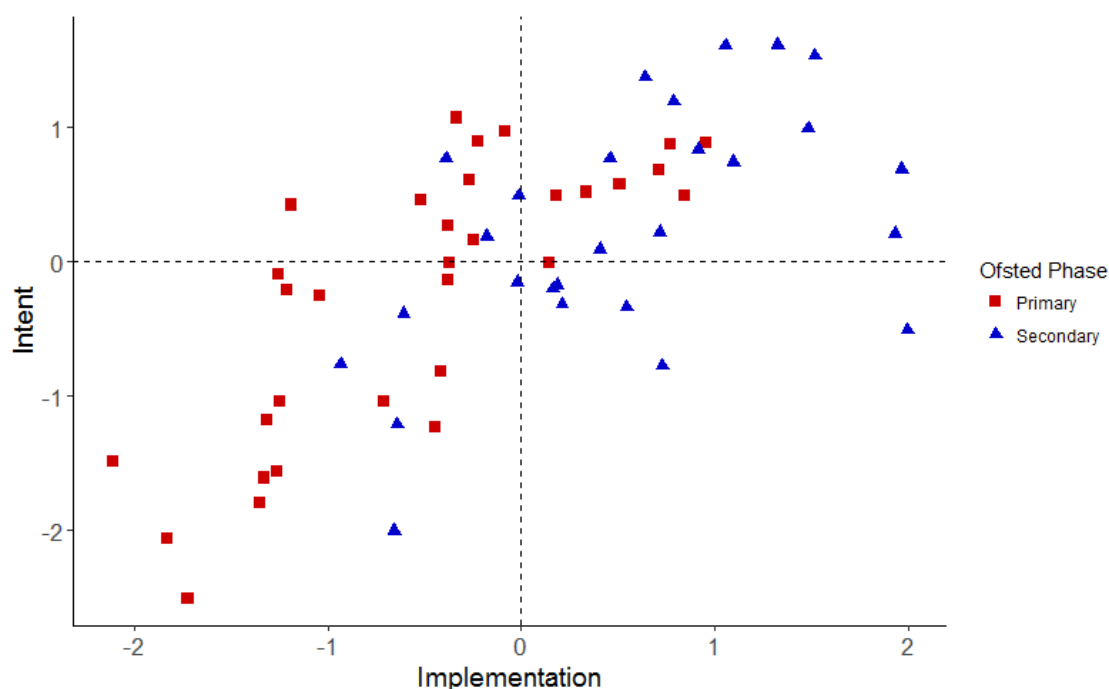
For instance, the top left quadrant shows approximately 10 schools that inspectors scored highly on their intent indicators but lower on the implementation indicators across the four subjects reviewed. This demonstrates that inspectors were able to establish weaknesses in curriculum implementation, despite the intended curriculum design being deemed as fairly strong. The distribution of the data also indicates this was a more common occurrence in the primary schools visited.

The evidence collected during the visits suggests that the mismatch in these schools was often down to leaders' knowledge of their curriculum being based on a perception of what is planned. In other words, leaders' accountability structures often checked the intended action of subject leadership, but too few of these leaders actually assured themselves that the planned curriculum was being implemented successfully. This point underscores that any evaluation of the curriculum on inspection must not be based on just conversations or a simple review of curriculum plans. The first-hand evidence of what pupils are taught is essential for constructing a valid overview of the quality of education.

Similarly, the data in the bottom right quadrant shows that the research model was capable of identifying schools (albeit only a few secondary schools in our sample) where leaders' intent for the curriculum was less convincing yet implementation remained strong. This reflects that curriculum expertise in some larger secondary schools exists within subject departments and not specifically with senior leaders. Figure 18 provides further details of the typical intent and implementation features that affected the schools found in each of the four quadrants.

Importantly, while the distribution of the data in figure 17 shows a correlation between the intent and implementation indicators, it also stresses the value of including subject leadership, multiple subject areas and a focus on first-hand evidence in the design process of the research model. For instance, it is likely that if these elements had not been included in the design, inspector observations for the top left and bottom right quadrants may have been missed. This would greatly affect inspectors' ability to report on the accuracy of overall curriculum quality.

The variation identified suggests, therefore, that our research model is valid. It appears to be assessing the right things, in the right way, to produce an accurate and useful assessment of curriculum intent and implementation.

Figure 18: Evidence of the trends found in schools when comparing the intent and implementation indicators from each quadrant in figure 17

Top left quadrant: Strong intent, weaker implementation of the curriculum	Top right quadrant: Strong intent, strong implementation of curriculum
■ Leaders focus on planning and paperwork but do not check its implementation or its impact ■ Subject leaders have complete autonomy. This goes unquestioned by the headteacher. ■ Subject leadership does not check the implementation of the curriculum and so the building blocks within units of work or schemes are not secure. This has an adverse impact on curricular implementation ■ Accountability (knowing what is implemented and learned) is narrowly focused on Year 2 and 6 in primary schools, and key stage 4 in secondary schools. ■ There are weaknesses in other non-benchmark years. These are not tackled in a timely way	■ High levels of accountability (knowing what is implemented and learned) ■ Clear methods to check what pupils know, can do and understand so that the right work is taught/informs teaching (assessment) ■ Teacher subject knowledge is consistently strong across the school, phase, key stage, and department ■ Senior leaders make it their business to check implementation of the curriculum – it is not left to chance ■ Leaders ensure that all groups of pupils can access the curriculum well ■ In primary schools, leaders understand all the component strands of the national curriculum – e.g. they check that the component parts of the technology curriculum happen – planning, designing, making and evaluating.
Bottom left quadrant: Weak intent, weak implementation of the curriculum	**Bottom right quadrant: Strong implementation and with weak leadership intent**
■ Accountability from the headteacher and subject leadership is poor (knowing what is implemented and learned) ■ Headteachers do not check implementation of the curriculum or delegate this task effectively. ■ There is a lack of accountability beyond English and mathematics ■ Accountability is about qualifications in the core subjects and data rather than the curriculum that is implemented and learned ■ Headteachers do not prioritise or know whether there are weaknesses in teacher subject knowledge ■ Little time or emphasis is given to subject leadership to check the impact of teaching ■ Progression across a key stage is weak ■ Units of work do not provide depth and this impedes pupils' conceptual understanding and subject specific knowledge over time.	Note: This was not a feature of any of the primary schools in the sample. ■ There is little strategic thought or decisions to shape the curriculum on offer beyond the teacher. ■ Weak intent by headteacher that impedes pupils' access to curriculum/an aspect of the curriculum ■ Teachers are left to deliver a curriculum. They have complete autonomy and the impact of teaching is consistently good, but the lack of coherence gets in the way of pupils' progression.

What are the most important indicators in the research model?

Although the two-factor model is a useful outcome of the factor analysis, it is worth noting that the factors and loadings produced above are less helpful in narrowing down the overall number of indicators to those that are the most important. This is due to the small sample size and high correlations between indicators.[8]

For instance, a correlation matrix showed that the 25 curriculum indicators, as well as the overall banding score, are very highly correlated with each other and statistically significant (p<=.01). This suggests that the indicators are very similar. This makes it difficult to refine them to a more manageable number because it is hard to distinguish between them and identify which are having the most impact. However, one interpretation of this is that most combinations of the indicators are likely to yield similar results, because they appear to be measuring very closely related aspects of the curriculum.

Additional correlations between the indicators and contextual factors were also investigated. Interestingly, none of the indicators, including overall banding, significantly correlated with IDACI quintiles. Ten of the indicators significantly correlated with the latest overall effectiveness grade of the schools in the sample. This suggests that the indicators are relatively independent of both a school's overall effectiveness grade and IDACI quintile. The correlation coefficients from this analysis can be found in Annex C.

In order to tease out the relative importance of the curriculum indicators in terms of predicting a school's overall banding score, we applied backwards regression modelling on the data to produce Model 1.[9] This technique indicates that the best model explains 99% of the variance between schools on the overall banding score and includes:

- one intent indicator (**curriculum is ambitious**)

- six implementation indicators (**subject leadership**, **subject knowledge**, **curriculum planning**, **equitable delivery**, **depth and coverage** and **progression model**)

- two impact indicators (**pupils successfully learn the curriculum** and **parity for pupils**).

---

[8] The model scores .84, in terms of the Tucker-Lewis Index, which is below the generally accepted criteria of >=.95, and scores .12 for the Root Mean Square Error of Approximation index, which is above the suggested criterion of <=.06 (Hu and Bentler, 1999).

[9] Backwards regression analysis looks to produce the model which has the best Akaike Information Criterion (AIC) score. The AIC is a relative measure of the quality of a predictive model, and a lower score indicates a better fit. This technique starts off with all 25 indicators included within the model, and then removes the indicator which reduces the AIC score by the largest margin. The process is repeated, until the model results in the lowest AIC score.

However, a caveat with Model 1 is that the high correlations between the indicators will make the backwards selection less clear and somewhat more random than if the correlations were lower. As high multi-collinearity appeared to be affecting the results from the previous run of factor analysis (figure 16), which included a large number of indicators in the factor loadings, this process was repeated but only on the indicators identified as important predictors within Model 1. Figure 19 demonstrates that factor 1 seems to cover indicators related to implementation, while factor 2 seems to cover indicators relating to intent. This replicates the findings from the original factor analysis, again correctly grouping the indicators to the intent and implementation design.

Figure 19: Results of the factor analysis conducted on Model 1 indicators
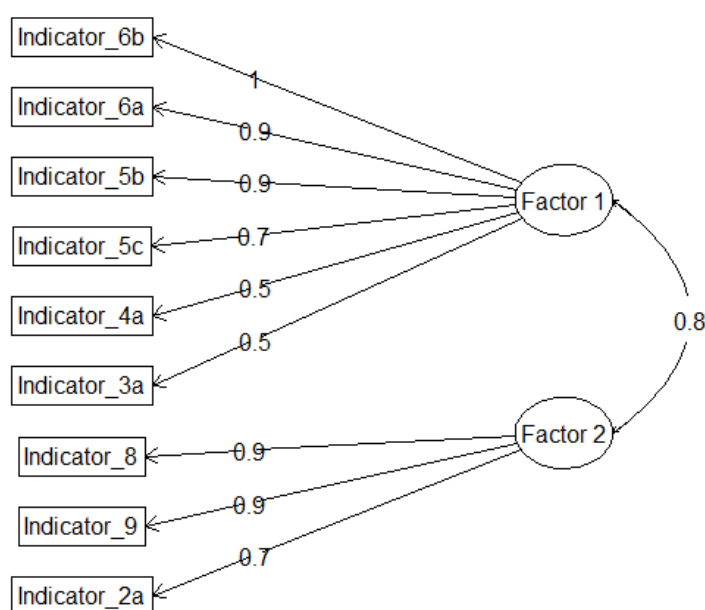
The two impact indicators, in some ways, can also be seen as summary indicators of the intent and implementation indicators themselves, much like the overall banding score. Figure 20 shows that they are very highly correlated with the overall banding:

Fig 20: Percentage of schools achieving each impact indicator rating, compared to the overall banding

	Band 1	Band 2	Band 3	Band 4	Band 5
Indicator 8	7	21	36	34	2
Indicator 9	2	20	36	36	7
Overall banding	5	25	33	36	2
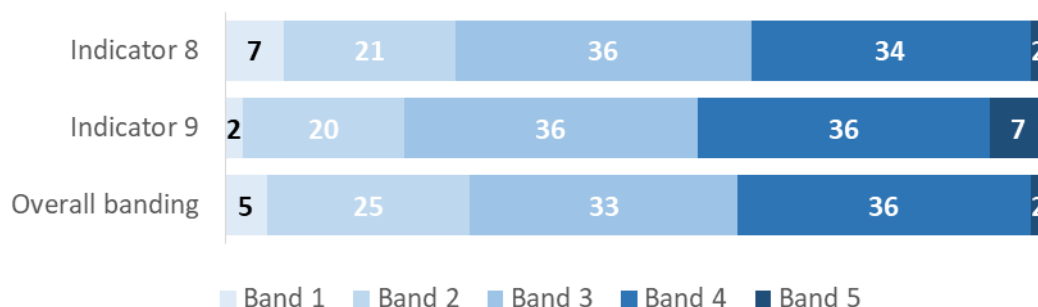
Figures are rounded and may not add to 100.

To remove their effect on the remaining indicators, analysis on Model 2 focused only on the intent and implementation indicators. Using backwards regression, as in Model 1, this produced a slightly different model to the previous regression analysis. The final model explained 97% of variance in a school's overall banding score and includes:

- three intent indicators (coherent rationale, curriculum is ambitious and centrally prescribed aims)
- seven implementation indicators (subject knowledge, regular curriculum review, curriculum planning, equitable delivery, progression model and coherence between intent and implementation).

While the statistical analysis is not as clear-cut as we had hoped in terms of narrowing down the indicators, we can still infer that indicators 2a (**curriculum is ambitious**), 3b and 4a (**subject knowledge**), 5b (**curriculum planning**), 5c (**equitable delivery**) and 6b (**progression model**) appear to be fairly important in determining a school's overall banding score and, therefore, overall curriculum quality. This is because they each appear in both Models 1 and 2. Further regression analysis is detailed in Annex D.

We can also infer from the models that the regular inclusion of implementation indicators on aspects of subject leadership, a curriculum progression model and curriculum planning may be the necessary focus for future indicators, rather than breaking them down into several indicators on a similar aspect. The two impact indicators also appear to be important, but their high correlation with the overall banding rating provided by inspectors suggests that they may actually be measuring the same thing (that is, overall curriculum quality).

What are inspectors' views on the most important indicators?

Following the visits, we asked the HMI involved in the fieldwork for their views on the process. In particular, we wanted to know which of the indicators they felt were the most important to draw on when it came to finalising their overall curriculum quality score. This was so that we could corroborate their perceptions with the statistical analysis we had completed to ensure that the right factors were included in our final model of curriculum quality.

In terms of importance, **coherent rationale**, **knowledge of curriculum concepts** and **curriculum is ambitious** were the most prominent intent indicators spoken about by inspectors during the focus group. They explained that the questions around a coherent rationale were useful because they prevented senior leaders from focusing the discussion too much on practical aspects of the curriculum, such as timetabling. This meant that the discussion opened up considerably, allowing inspectors to identify variation between the schools.

Some leaders were 'floored' by the focus on a coherent rationale, because it seemed quite an unexpected question. However, this was often related to those leaders struggling with curriculum design and their ability to respond effectively and explain the purpose of their curriculum.

Knowledge of curriculum concepts was discussed in a similar vein and appeared useful for picking up on strengths and weaknesses of the posited curriculum intent, for instance where leaders were confusing an educational intervention as a curriculum theory or applying research to their design but had developed no means for evaluating its impact on pupils.

One point inspectors raised was the number of subjects that would need to be seen in order to effectively assess the whole curriculum. For instance, a couple of inspectors indicated that the larger size of secondary schools, and the number of specialisms they may offer, could make it more difficult for inspectors to 'get a grip' on the curriculum during inspection.

HMI discussed each implementation indicator in depth, but generally advocated that **subject leadership knowledge** was essential. The focus on the **progression model** was mentioned by all the participants as being helpful to explain whether leaders were able to build a curriculum or had a concept of curriculum progression. Inspectors also felt that **assessment of the curriculum** was integral to the success of the progression model, particularly for ensuring that they were able to identify gaps in knowledge or misconceptions in pupils' learning.

What do the evidence forms tell us about inspector reliability?

As part of the post-visit analysis, we reviewed the evidence forms to check that the data quality was sufficiently valid – that is, to what extent the evidence relates to the question prompts and content of the rubric so that the ratings given on the five-point scale can be considered accurate. A mismatch between the text in the evidence base and the quantitative score given could indicate a degree of unreliability in the subjective assessment made by an inspector.

An overview of the findings from this analysis can be found below.

■ In all the evidence forms, HMI collected first-hand evidence about curriculum implementation. This evidence often led to pertinent discussions with leaders about the impact of the school's curriculum offer. As a result, inspectors were able to make accurate judgements about the quality of the curriculum on offer.

■ When the match between the text in the evidence forms and the quantitative score given was consistently strong, inspectors had followed the essential features of the research methodology. For instance, they were able to evaluate curriculum implementation, when they had examined first-hand evidence and real examples of pupils' work, quickly. The evidence forms also captured productive conversations with subject leadership about the relationship between the planned curriculum and the impact of the curriculum on offer.

■ In a few cases, however, inspectors tended to record what leaders 'said' or the 'processes in place' more than looking first hand at the 'depth, coverage and impact' of curriculum implementation. In these instances, the text to score match was weaker because evidence was based more on dialogue

with leaders and their perception of curriculum and less time was spent working alongside leaders to evaluate how they had implemented the curriculum.

- On a few occasions, recorded evidence focused on the processes rather than the impact that the curriculum was having on pupils' learning. The evidence forms therefore did not adhere to the implementation indicators and rubric well enough. This suggests that future inspection of the curriculum should not be limited to a conversation about leadership processes in designing a curriculum.

- Evidence collected for different curriculum indicators did vary. However, this owed more to whether subject leaders had considered a specific indicator (such as developing a progress model or providing depth and breadth to the content) within their curriculum design than inspector differences.

- A few evidence bases did not depict specific enough information about the indictors or the impact of the curriculum offer. Most of the evidence forms, however, were consistently detailed and provided sufficient evidence that linked to the research method. This ensured an appropriate match between text and indicator scores.

The evidence forms suggest that, generally, the rubric was applied by inspectors in a reliable way. Although in rare cases the score on the rubric did not quite match the text of the evidence forms, this was typically only for a couple of indicators per evidence base.

In addition, the quality of the evidence bases improved as more visits were completed. This suggests that the quality assurance mechanism, where the HMI leads checked the first evidence base of other HMI colleagues to provide feedback, worked as we intended. It also suggests that inspector training is paramount, quality assurance is required to ensure consistency and monitoring of evidence bases and outcomes to confirm text to score match is essential. As a result, we will need to consider these aspects carefully when the new framework is established.

What did the evidence tell us about the curriculum in schools?

Curriculum leadership

The evidence showed that senior leaders in the schools that were assessed as 4 or 5 for curriculum quality ensured that the planned curriculum was implemented across a wide range of subjects. Typically, this meant devolving curriculum decisions to others with curriculum and subject expertise, but holding subject leadership to account for checking the coverage and the depth of knowledge that pupils learn. This often involved:

- assuring themselves that middle leaders with responsibility for leading subjects have the right skill set and subject knowledge to carry out their roles effectively

- providing high-quality professional development to develop teacher subject knowledge beyond the core subjects; in the primary schools, leaders were often facilitating development through the subject associations and internal information-sharing if there was an expert in the school

- checking the delivery and impact of the curriculum as an instrumental part of whole school improvement

- ensuring that delivery does not focus narrowly on tests and qualifications and valuing all subjects so that pupils receive equitable provision across a range of subjects

- ensuring that teaching in foundation subjects was not reduced to time-filling exercises – or 'down time' from the core subjects – that do not develop pupils' conceptual understanding of subject disciplines.

The subject leaders in these schools often showed a thorough understanding that progress is through learning the curriculum. They tended to hold onto strong beliefs around teaching content that they felt best met their pupils' needs and were keen to avoid teaching the curriculum in a reductive way. Senior leaders supported this by recognising that individual subjects required different approaches to sequencing and assessing content. They generally avoided a one-size-fits-all approach to curriculum development, supported by strong subject leadership that understood their subject content well. Curriculum accountability was particularly strong in these schools because leaders took determined action to sort out any deficiencies in curriculum provision in a timely way, once any shortcomings were identified. Strong principles around assessment linked to the curriculum progression model ensured that these schools could quickly address gaps in pupils' knowledge and understanding.

Special schools and other school types

The visits to the two special schools in the sample showed that the curriculum indicators worked just as well. Both of the inspectors who carried out these visits (who were SEND specialists) confirmed that the intent and implementation design was as effective in drawing out the first-hand evidence as it was for the other schools they had visited.

Inspectors found that the research model worked well across a number of different school types, for example, infant and junior schools, academies, small rural schools and the two special schools. One area of interest was around how inspectors determine the curriculum intent of a school in a multi-academy trust when the curriculum they are delivering has actually been designed by a central hub school within the trust. The research suggests that it is important to speak with trust leaders about the intended and implemented curriculum as well as the individual academy. It will be important for inspectors to include conversations with trust staff when unpicking the curriculum offer and establishing the quality for its pupils in the new framework.

Key stage 2 narrowing

As shown by the descriptive statistics on page 13, too many primary schools in our sample had an imbalanced curriculum offer. In these schools, the curriculums were not as challenging as that set out in the national curriculum 2014. In particular, the foundation subjects were less well implemented than core subjects, although in some cases the curriculum intent of leaders was also largely absent.

Inspectors identified that even where overall curriculum quality was weak, the English and mathematics curriculums were typically well planned in terms of the knowledge, skills and understanding they wanted their pupils to develop. It was clear that sequencing and progression of concepts were often being considered. Pupils' work showed this as well. These features were also reflected in other subject areas in the primary schools that scored band 4 or 5. The leaders of these schools tended to view subjects as individual disciplines, even when they were being delivered as part of a topic or thematic approach. This meant there was no trade-off between strong performance in foundation subjects and core subjects.

By contrast, the weaker primary schools paid limited attention to the knowledge, skills and understanding required in the foundation subjects. This was sometimes because the subject lens was lost through the use of a topic-based approach. In other circumstances, there was a lack of subject knowledge among staff to develop an enriching offer for certain subjects. Furthermore, there were a few schools that were focused on content and had a curriculum plan in place, yet the planning was only at a surface level and the content being delivered was piecemeal, with the objective solely of ticking off a box on the national curriculum check-list.

In these cases, pupils were not able to develop knowledge of the subject discipline sufficiently well. Inspectors noted that the work carried out by pupils in the weakest primary schools visited was often undemanding, with pupils completing low-level tasks such as ordering or gluing and sticking. There were also several instances when progression in foundation subjects was not secure because pupils' outcomes were assessed against writing criteria. In other words, the subject-specific knowledge that was intended to be taught was lost among generic writing criteria. Some foundation leaders also reported that senior leaders promoted this and that written outcomes in history and geography were vehicles to successful writing assessments.

Key stage 3 narrowing

Around half of the secondary schools visited were continuing to deliver a three-year key stage 3 model. The reasons for this included ensuring that pupils experience a wide curriculum in order to prepare well for key stage 4 and pupils' increased maturity and knowledge to make well-reasoned decisions about their future studies. The leaders of the schools that operated a two-year key stage 3 model or a 'mixed' model (where some subjects begin teaching key stage 4 requirements in Year 9) cited 'minimising the wasted years' or the need to lengthen key stage 4 to meet the demands of the new exam-board specifications as rationales. This is despite the new GCSEs being designed to have content lasting two years.

Some leaders were taking steps to improve uptake of the EBacc. However, this remained low and deemed to not be a priority by leaders and governors in a considerable number of the schools visited. Some schools deemed the EBacc as not being appropriate for the context of the school and its pupils in terms of prior attainment or destination outcomes. In some cases, leaders had already decided what certain groups of their Year 7 cohort would be capable of – not understanding that progression is not linear – and had constrained the curriculum offer. This means that pupils who may have a talent for a particular subject were not being given suitable access to expose those talents.

The timetabling and the organisation of curriculum delivery in some of the schools with weaker curriculum quality also limited pupils' knowledge and understanding in technology and arts subjects in key stage 3. Practical and creative subjects were sometimes marginalised. It is important, however, to note that headteachers in the schools with a band 4 or 5 for curriculum quality were often passionate advocates of the benefits of subjects such as music, drama and technology. A wide range of subjects tended to thrive in these schools.

Curriculum equity

Leaders commonly talked about making the curriculum appropriate to the context of their school. This reflects the findings from phase 2 about this being a main aim for many leaders. However, in some schools, this was not always implemented effectively. For instance, equality of opportunity was often an issue for lower attaining pupils or pupils with special educational needs and or disabilities (SEND). The strategy employed by leaders for these pupils sometimes involved teaching primarily by non-specialists (in lower sets) and teaching assistants. Evidence from the Education Endowment Fund suggests that teaching assistants being used to substitute rather than supplement teaching from teachers can have negative consequences for pupils' outcomes.[10]

Equity was also an issue in schools where English or mathematics catch-up for some pupils was timetabled in a way that meant that they missed out on other curriculum opportunities that were otherwise available to other pupils. This was picked up in secondary schools as an issue where lower ability pupils were, typically, taken out of modern foreign languages lessons in key stage 3 so that they could receive further lessons in English or mathematics. However, the more effective schools appeared to have curriculum strategies in place that allowed both equity and achievement to be attained for such pupils. This typically involved small-group teaching covering the same content as other pupils, rather than using catch-up interventions.

Importantly, getting the balance right between the tensions of a fair curricular offer and providing support for pupils who need it so they can access that offer is a

---

[10] Teaching assistants; Education Endowment Fund, 2018; https://educationendowmentfoundation.org.uk/evidence-summaries/teaching-learning-toolkit/teaching-assistants/

challenge. Unfortunately, in a few primary schools, despite there being a broadly equal access to the curriculum in place, occasionally, a 'lost curriculum' was identified. This occurred when the intent was for all pupils to access the same curriculum, but in reality some teachers did not teach the full range of key ideas and teaching was not sufficiently challenging. This lowering of expectations was typically associated with science and had a more detrimental effect on the higher ability learners. When this was done on an 'ad-hoc' basis, with little or no leadership oversight, senior leaders were unaware of when and where the school's curriculum offer was weak. As a result, this had a detrimental effect on the pupils' knowledge, skills and understanding over time.

Subject knowledge

The evidence bases for both primary and secondary schools showed that strong teacher subject knowledge is essential to high-quality curriculum planning. Subject knowledge ensured both the depth of coverage and an understanding of content sequencing that related to the subject discipline. Strong subject knowledge often resulted in a higher overall curriculum quality score, because its impact on curriculum design and pupils' knowledge, skills and understanding was particularly noticeable by inspectors.

The leaders of the secondary schools in bands 4 and 5 for curriculum quality had tended to prioritise recruiting staff with specific subject knowledge. Alongside pursuing a full complement of subject experts within departments, these leaders had also thought about retention through curriculum design. In these cases, staff were given greater control of and were directly involved in curriculum decision-making through trusting relationships with senior staff. They used their expertise to design the curriculum, rather than having the content dictated to them by leaders. Importantly, the better schools still had an accountability structure in place that ensured that leaders had oversight of their experts' curriculum decisions.

Of course, with current teacher recruitment and retention issues it is difficult for all secondary schools to feature a full complement of subject specialists on roll.[11] Furthermore, staff in primary schools are not all going to have expertise in all areas of the national curriculum. However, leaders in some schools were doing little to embed or develop their staff's subject knowledge to improve learning. For instance, professional development in English or mathematics, through external courses or working with local hubs and other schools to share ideas, was common place. Yet opportunities for teaching staff to receive professional development in the foundation subjects was infrequent. On occasions, subject leaders in secondary schools and some primary schools were driving this for subject colleagues through their links with subject associations, but it was not something being expressed routinely as a priority by senior leaders.

---

[11] Worth, J., Lynch, S., Hillary, J., Rennie, C. and Andrade, J. (2018). Teacher Workforce Dynamics in England. Slough: NFER.

In some of the primary schools, a number of leaders and staff were unsure about who to contact to provide subject-specific professional development. In a few cases, this was because local authority subject teams had been disbanded; in others, it was because subject staff had limited time to research and pinpoint the additional professional development that was required. As a result, subject leadership lacked the external support required to develop their subjects. This restricted the depth and coverage of the curriculum on offer. In the primary schools in bands 1 and 2 for curriculum quality, leaders typically did not prioritise subject specific professional development.

Curriculum planning

The discussions with leaders about curriculum intent suggested that many had a good understanding of their context and had clear aims for all pupils to progress. However, in the schools with weaker overall curriculum quality, headteachers were less able to articulate how to map and build progression, because they were unsure what it looked like for their school.

In the weakest schools, particularly in primary schools, this tended to mean having a set of curriculum objectives or subject topics available as part of a planning document. However, the first-hand evidence revealed that this documentation was often designed only at a superficial level. Inspectors saw that these leaders were frequently not considering sequencing between components of knowledge that would lead to conceptual understanding. As a result, the curriculum design in these schools was focused on delivering content, but with a lack of consideration for the ordering and structure of that content.

In some cases, curriculum design was little more than cutting and pasting the key objectives of the national curriculum. In other words, they merely highlighted the key statements in the national curriculum and planned around those. A focus on what had been taught and learned in these schools was therefore limited. The work scrutiny evidence offered little assurances that anything of substance had been learned effectively. All too often, teaching was being driven by specifc activities rather than ensuring that activities were delivering an ambitious curriculum. This evidence suggests, therefore, that the future inspection focus on curriculum cannot use documentation and planning documents alone to understand the extent of curriculum implementation.

Assessment

The evidence forms for the secondary schools visited identified some weaknesses in how curriculum assessment was being applied. A common finding for the secondary schools in band 2 or 3 for curriculum quality was that they were assessing pupils in key stage 3 using GCSE grades 1 to 9 from the beginning of Year 7. This is problematic as this type of assessment is not focused on the key stage 3 curriculum being delivered and what pupils actually knew and understood.

Furthermore, evidence from the HMI focus group showed that in some schools a homogeneity of approach was found across all subjects. In these cases, leaders had

established a whole-school assessment approach that used exam models to structure their notions of progress, but had not considered what progression looked like in individual subjects. This indicated that the specific progression of knowledge, including disciplinary knowledge of the subject, was not being considered by leaders. However, it was commonly mentioned by senior leaders and subject leads in these schools that assessment continued to be a significant piece of work for them and that it was not yet fully fit for purpose or that it was still 'a work in progress'. A few, though, did not see the whole-school approach to curriculum assessment as being a problem.

Transition

The variation in quality of primary school curriculums has some implications for how secondary schools manage the curriculum offer in Year 7. As the data shows, secondary schools generally have a wider curriculum than primary schools. However, a common theme from speaking to subject leads was the wide variance in knowledge of Year 7 pupils on transition to secondary schools, mostly in the foundation subjects but occasionally in English and mathematics too. This often made it difficult for schools to asses what pupils already knew, particularly if a secondary school had multiple primary feeder schools. Subject leads frequently told us that this was a challenge. Transition remains a particularly tricky road bump for secondary schools and has serious consequences for curriculum progression if not managed appropriately.

Next steps

The study reveals that our research model has good face validity. The evidence points towards a structure and process that works in different school contexts and does not favour one curriculum approach over another. Importantly, the intent and implementation design seems to allow inspectors to distinguish between effective and ineffective curriculum practice, contributing to an accurate picture of the quality of education being provided for pupils. The evidence also provides details of the main predictors of curriculum quality that may benefit a more refined model for assessing the quality of education in the future. Overall, the findings provide a degree of confidence that our plans to look beyond data and assess the broader quality of education are achievable and necessary.

The focus of this study was to provide a research basis that will feed into the design of the new framework. The next steps will see features of the model trialled in pilot inspections. This will allow us to assess how the indicators and the processes for getting at first-hand evidence work under inspection conditions. Furthermore, as the curriculum research has shown, some challenges remain, particularly for how we calibrate our judgement profile in view of the primary school findings. Our aim is to consult with the sector over the spring term to take on board further views on how the quality of education should be inspected in the future.

Annex A: Detailed visit process methodology

Initial meeting with senior leaders

Phase 2 of our study highlighted that the discussion with curriculum experts was useful in order to understand curriculum intent. We replicated this process within the research visits for phase 3, although we reduced the time from two hours to between 30 and 45 minutes. This reduction meant that inspectors had more time to gather evidence that could underpin the successful implementation of leaders' curricular intent.

The main purpose of the initial meeting was to collect evidence that aligned with the curriculum intent indicators (indicators 1a to 2d). We could then triangulate this information with the first-hand evidence collected across the other research activities to better understand the relationship between intent and implementation. A secondary purpose was to trial whether this method was suitable for an actual inspection.

The questions were adapted to meet this timescale and allowed school leaders to explain the organisation and structure of their curriculum offer. Leaders in primary schools, for example, could talk about how the thematic curriculum was organised or whether specific subjects were taught by one teacher. The questions also allowed senior leaders to share the 'big picture' of the school's curriculum, its aims and the rationale that sit behind it.

Subject leaders' meeting

The main part of the research visit involved meeting with subject leaders. The purpose of these sessions was to gather first-hand evidence about implementation and impact of the curriculum. The activities involved allowed inspectors to assess how well:

- leaders assure themselves that the school's curriculum intentions are met and the subject curriculum is sufficiently challenging for the pupils it serves

- the content, sequencing and progression in the subject is appropriate and secured and whether it demands enough of pupils

- all domains/strands within the subject are learned (as set out in the national curriculum)

- pupils consolidate their knowledge, understanding and skills in the subject

- the school's curriculum offer prepares pupils for their next stage.

On each visit, inspectors led four subject meetings across the day. Each meeting was scheduled to take 50 minutes. Subjects were grouped so that the inspector was able to review a range of subject types, for example a core subject, a STEM (science, technology, engineering and mathematics) subject, the humanities and the arts. The process differed slightly between phases. For instance, in the secondary schools visited, these meetings often occurred in the subject department area or

'base' so that inspectors could more easily view other evidence related to the subject, such as a short observation of a lesson.

The focus of these meetings was on subject leaders' intent of the curriculum – in particular, any deliberate actions they had taken over the last 18 months to shape the curriculum offer through the lens of the subject – and working alongside the subject lead to review the implementation of the subject. Importantly, we established that these sessions could not just be a discussion with the subject lead. Inspectors had to organise the meeting so that it focused on reviewing first-hand evidence. One thing we advised inspectors not to look at as part of this design was internal school data on pupils' achievement. Figure 21 shows the standardised process developed for this part of the study design.

This approach allowed inspectors to look at work in books, portfolios or other pupil outputs and compare it with any written documentation that already existed in the school to corroborate the curriculum plan against the curricular intent. We designed the questions asked by inspectors, particularly on how the progression of pupils' learning was checked across years and whether teaching staff had received recent training in their subject specialism, to support this critical assessment of curriculum implementation.

Meeting with pupils

We also convened a meeting with pupils so that we could triangulate their views with the evidence collected from the subject lead sessions. The focus of the pupil discussion was related to the outcomes from the subject lead meetings and required pupils to elaborate on:

- how often the subjects of the visit were taught
- the design and implementation of the curriculum
- how motivated pupils are about these subjects
- pupils' retention of what had been taught in these subjects
- pupils' knowledge of domains and strands within a subject discipline
- additional questioning about the coverage of other subjects not part of the subject foci of the visit
- pupils' views about their transition from class to class and between key stages.

Figure 21: Process applied during the subject review with subject leadership

Intent Talk to me about the **deliberate action you have taken to shape the curriculum**	Talk to me about timetabling, coverage, mapping of subject	**Gathering first-hand evidence of implementation** Work through pupils books to demonstrate **progression across a key stage**. E.g. take a strand of that subject. Ask leaders to show you how this plays out in the books or pupils' output Short learning walk around the department, including speaking to pupils and teaching staff about stretch and challenge and knowledge coverage Curriculum mapping and assessment materials to help identify extent of knowledge and skills coverage and depth in the curriculum	Show me how teachers assess across unit, term, year, key stage Record first-hand evidence	Impact of staff training Staff specialisms Staff subject knowledge Record first-hand evidence

Impact

During the book scrutiny you will be examining the impact of the implementation of the curriculum.
E.g. does the school's design of KS 3 performing arts curriculum/carousel activities limit pupils' exposure to music?
Does an absence of formative assessment systems impact on the curriculum quality?
LPA pupils are taught by non-specialists in all year groups – what impact does this have on the implementation and impact of the curriculum for these pupils?
Do all groups have access to the curriculum? How do you know?
Are all staff in their department subject specialists? What impact does this have on curriculum implementation?

Initial discussion   Collection of first-hand evidence

In secondary schools, the discussion was extended to cover the following aspects:

- the process for choosing key stage 4 option subjects
- how well prepared pupils feel for their next stage/destination
- what the school has done to prepare pupils for their next stages
- organisation across subjects, e.g. carousels and how tutor time is used
- subject content – progression, gaps, coverage across key stages.

The aim was to speak to between four and 12 pupils during the pupil discussion. Inspectors selected the pupils on the basis of the work books they had reviewed during the subject lead meetings. The pupil discussions were planned to take no longer than 20 minutes. Owing to the constraints of a one-day visit, particularly in attempting to provide coverage across four subject areas, only one or two pupil group discussions were scheduled per visit.

Meeting with governors

We also met with school governors to understand their role in the strategic decisions that had been made to implement the school's curriculum offer. This generally involved a short face-to-face meeting at the end of the visit, although in a few cases we scheduled a short telephone discussion with a representative of the governing body when none were available to attend the face-to-face discussion. Standardised questions were prepared to ensure consistency across the visits. In particular, this discussion focused on how well governors were informed about the success of the curriculum in preparing pupils for their next stage.

HMI focus groups

Following the completion of the fieldwork, five of the inspectors participated in a focus group to feed back on the research model and visit process. This was led by the senior research lead for the project. HMI availability meant that the other four inspectors were unable to join the focus group, although some of their views were picked up through correspondence. The main questions asked were around identifying which aspects of the research design the inspectors thought worked well and where they had concerns around implementation of the research model, particularly in relation to future inspection practice. This evidence was used to supplement the data analysis in refining and narrowing the indicators to those that are most important for assessing curriculum quality.

Protocols

This project went through an ethical approval process which followed British Education Research Association (BERA) guidelines. The school leaders who were contacted about a visit were made aware that this would be part of a research study. We explained to them that this would differ from a routine inspection and that the information collected as part of the study would not be shared with Ofsted's regions to inform any future inspection of the provision. We shared details of the

An investigation into how to assess the quality of education through curriculum intent,
implementation and impact
December 2018, No. 180035

aims and methods of the project in advance. Visits only took place with the consent of school leaders.

For this study, we decided that inspectors would not provide feedback to leaders about the quality of their schools' curriculum. As the purpose of this research was to determine the validity of the research model in assessing curriculum quality, we felt that it was inappropriate to share the visit findings on what were, at that point, untested indicators. The data-tables that follow in the evaluation section of this report, therefore, have some statistical disclosure methods applied. This is so inferences cannot be made against the characteristics of individual schools. A feedback session took place at the end of the research visit, to enable leaders to feed back to inspectors their views and experience of the process.

Inspectors also reinforced that they did not have a preconceived view of curriculum, except that there should be progression in what pupils know, can do and understand over time. The impact of the school's actions to implement the curriculum, evaluated against the key criteria and questions from the research model, were clarified as the core purpose of the research. Inspectors also made it clear that they did not require information to be presented in a particular format and that there was no need to prepare any new material ahead of the visit.

Annex B: Sample tables

Fig 22: Sample breakdown by latest overall effectiveness judgement at time of selection

Overall effectiveness	Total no. of schools	Primary	Secondary
Outstanding	19	9	10
Good	30	17	12
Requires improvement	15	7	7
Total	64	33	29

One of the special schools visited was judged good for overall effectiveness at its last inspection, the other was judged requires improvement.

Fig 23: Sample breakdown by Ofsted region

Ofsted region	Total no. of schools
East Midlands	7
East of England	10
London	9
North East, Yorkshire and the Humber	8
North West	7
South East	6
South West	10
West Midlands	7
Total	64

Fig 24: Sample breakdown by IDACI quintile

IDACI	Total no. of schools	Primary	Secondary
Quintile 1 (least deprived)	13	7	6
Quintile 2	14	7	7
Quintile 3	11	5	6
Quintile 4	13	8	4
Quintile 5 (most deprived)	13	6	6
Total	64	33	29

One of the special schools visited was located in IDACI quintile 4, the other was in quintile 5.

Fig 25: Sample breakdown by school type

School type	Total no. of schools
Primary	
Academy Converter	10
Community School	13
Foundation School	4
Free School	1
Voluntary Aided School	3
Voluntary Controlled School	2
Total	33
Secondary	
Academy Converter	12
Academy Sponsor Led	2
Community School	4
Free School	5
Voluntary Aided School	6
Total	29
Grand total	62

Annex C: Correlations between quality indicators, IDACI quintile and overall effectiveness grade

Variables	OE	IDACI
Indicator 1a	-.11	-.02
Indicator 1b	-.09	.01
Indicator 1c	-.27*	.05
Indicator 1d	-.24	-.10
Indicator 2a	-.14	-.08
Indicator 2b	-.22	-.01
Indicator 2c	-.23	-.03
Indicator 2d	-.23	.03
Indicator 3a	-.30*	-.12
Indicator 3b	-.28*	-.07
Indicator 3c	-.26*	.01
Indicator 4a	-.31*	-.09
Indicator 4b	-.25*	.03
Indicator 5a	-.33**	-.07
Indicator 5b	-.22	.04
Indicator 5c	-.22	-.09
Indicator 5d	-.12	.04
Indicator 6a	-.17	-.02
Indicator 6b	-.17	.00
Indicator 6c	-.13	.00
Indicator 7a	-.25	.07
Indicator 7b	-.28*	.07
Indicator 7c	-.30*	.03
Indicator 8	-.25	.08
Indicator 9	-.27*	.04
Overall banding	-.22	.10

* p <= .05, ** p<= .01, *** p<= .001

Annex D: Additional regression analysis

Additional modelling techniques were also carried out on the indicator data collected to see whether they identified similar or different predictors on the school's overall banding score. Model 3 applied manual backwards regression to the same data in Model 1 to overcome the inclusion of some indicators in Model 1 that were not significantly contributing to its AIC score.[12] The indicator with the lowest p-value was removed at each step with the process being repeated until the model only included indicators that were statistically significant ($p<=.05$). The final model obtained from this technique explained 99% of the variance in a school's overall banding score, and included: **subject leadership**, **pupils successfully learn the curriculum** and **parity for pupils**.

Model 4 used forwards regression, which starts off with no variables in the model and adds in one indicator at a time to produce the best fit; and Model 5 stepwise regression, which works in both the forwards and backwards directions to produce the best fitting model. Both procedures resulted in the same final model that explained 99% of variance in a school's overall banding score, and contained **curriculum planning**, **pupils successfully learn the curriculum** and **parity for pupils**.

---

[12] The AIC is a relative measure of the quality of a predictive model, and a lower score indicates a better fit.

Annex E: Correlations between curriculum indicators and overall banding scores

Variables	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Indicator 1a	-												
2. Indicator 1b	.75***	-											
3. Indicator 1c	.74***	.81***	-										
4. Indicator 1d	.78***	.63***	.65***	-									
5. Indicator 2a	.78***	.65***	.58***	.76***	-								
6. Indicator 2b	.56***	.65***	.60***	.52***	.69***	-							
7. Indicator 2c	.60***	.67***	.67***	.55***	.52***	.53***	-						
8. Indicator 2d	.61***	.77***	.65***	.65***	.54***	.55***	.70***	-					
9. Indicator 3a	.64***	.59***	.59***	.66***	.65***	.55***	.49***	.51***	-				
10. Indicator 3b	.61***	.55***	.63***	.63***	.70***	.50***	.49***	.45***	.85***	-			
11. Indicator 3c	.47***	.46***	.48***	.61***	.58***	.41**	.51***	.46***	.80***	.74***	-		
12. Indicator 4a	.58***	.54***	.54***	.67***	.67***	.54***	.52***	.52***	.80***	.80***	.78***	-	
13. Indicator 4b	.57***	.50***	.48***	.71***	.67***	.46***	.44***	.47***	.79***	.75***	.83***	.87***	-
14. Indicator 5a	.53***	.47***	.55***	.67***	.63***	.64***	.46***	.50***	.78***	.75***	.75***	.85***	.76***
15. Indicator 5b	.50***	.50***	.49***	.59***	.59***	.55***	.44***	.47***	.75***	.67***	.69***	.70***	.68***
16. Indicator 5c	.57***	.48***	.44***	.64***	.63***	.49***	.48***	.46***	.73***	.63***	.68***	.69***	.69***
17. Indicator 5d	.53***	.48***	.47***	.51***	.67***	.46***	.50***	.43***	.65***	.67***	.71***	.69***	.68***
18. Indicator 6a	.45***	.36**	.41**	.51***	.61***	.51***	.36**	.34**	.64***	.68***	.66***	.73***	.68***
19. Indicator 6b	.50***	.45***	.54***	.57***	.62***	.57***	.44***	.41***	.67***	.71***	.69***	.76***	.68***
20. Indicator 6c	.53***	.44***	.49***	.58***	.63***	.46***	.37**	.36**	.72***	.72***	.73***	.76***	.74***
21. Indicator 7a	.48***	.50***	.57***	.54***	.64***	.54***	.50***	.46***	.68***	.77***	.70***	.79***	.75***
22. Indicator 7b	.53***	.54***	.63***	.57***	.66***	.53***	.53***	.49***	.69***	.78***	.69***	.77***	.74***
23. Indicator 7c	.55***	.47***	.58***	.63***	.62***	.55***	.44***	.47***	.75***	.80***	.69***	.82***	.77***
24. Indicator 8	.60***	.56***	.56***	.67***	.69***	.65***	.46***	.48***	.68***	.64***	.67***	.73***	.75***
25. Indicator 9	.59***	.51***	.51***	.67***	.62***	.69***	.43***	.49***	.63***	.53***	.61***	.68***	.70***
26. Overall banding	.62***	.56***	.55***	.65***	.70***	.66***	.45***	.50***	.70***	.63***	.66***	.69***	.71***

* p <= .05, ** p<= .01, *** p<= .001

Variables	14	15	16	17	18	19	20	21	22	23	24	25
1. Indicator 1a												
2. Indicator 1b												
3. Indicator 1c												
4. Indicator 1d												
5. Indicator 2a												
6. Indicator 2b												
7. Indicator 2c												
8. Indicator 2d												
9. Indicator 3a												
10. Indicator 3b												
11. Indicator 3c												
12. Indicator 4a												
13. Indicator 4b												
14. Indicator 5a	-											
15. Indicator 5b	.74***	-										
16. Indicator 5c	.74***	.79***	-									
17. Indicator 5d	.67***	.74***	.78***	-								
18. Indicator 6a	.79***	.79***	.70***	.68***	-							
19. Indicator 6b	.80***	.84***	.73***	.77***	.88***	-						
20. Indicator 6c	.75***	.79***	.72***	.71***	.87***	.88***	-					
21. Indicator 7a	.78***	.75***	.69***	.76***	.79***	.85***	.81***	-				
22. Indicator 7b	.77***	.73***	.66***	.75***	.76***	.81***	.76***	.95***	-			
23. Indicator 7c	.86***	.80***	.74***	.67***	.79***	.83***	.80***	.85***	.85***	-		
24. Indicator 8	.71***	.67***	.58***	.52***	.68***	.69***	.74***	.74***	.71***	.73***	-	
25. Indicator 9	.72***	.67***	.65***	.52***	.62***	.67***	.67***	.61***	.57***	.66***	.85***	-
26. Overall banding	.69***	.72***	.59***	.58***	.65***	.73***	.70***	.69***	.66***	.71***	.90***	.85***

* p <= .05, ** p<= .01, *** p<= .001

Annex F: List of the 64 schools that participated in the research visits

School name	Local authority	Type	Phase
Arnold Woodthorpe Infant School	Nottinghamshire	Community School	Primary
Babington Academy	Leicester	Academy Converter	Secondary
Birches Green Infant School	Birmingham	Community School	Primary
Broadwater Primary School	Wandsworth	Community School	Primary
Carville Primary School	North Tyneside	Foundation School	Primary
Castle Manor Academy	Suffolk	Academy Converter	Secondary
Chapelford Village Primary School	Warrington	Academy Converter	Primary
Chetwynde School	Cumbria	Free School	Secondary
Chingford CofE Primary School	Waltham Forest	Voluntary Controlled School	Primary
Churchmead Church of England (VA) School	Windsor and Maidenhead	Voluntary Aided School	Secondary
City Academy Birmingham	Birmingham	Free School	Secondary
Corsham Primary School	Wiltshire	Academy Converter	Primary
Cosgrove Village Primary School	Northamptonshire	Community School	Primary
Cowley International College	St Helens	Community School	Secondary
Crossley Hall Primary School	Bradford	Community School	Primary
Ditton Park Academy	Slough	Free School	Secondary
Earlsdon Primary School	Coventry	Community School	Primary
Eden Girls' School Coventry	Coventry	Free School	Secondary
Elmridge Primary School	Trafford	Academy Converter	Primary
Figheldean St Michael's Church of England Primary School	Wiltshire	Academy Converter	Primary
Filey Church of England Nursery and Infants Academy	North Yorkshire	Academy Converter	Primary
Fir Vale School	Sheffield	Academy Converter	Secondary
Fowey River Academy	Cornwall	Academy Sponsor Led	Secondary
Harris Girls Academy Bromley**	Bromley	Academy Converter	Secondary

School name	Local authority	Type	Phase
Holway Park Community Primary School	Somerset	Community School	Primary
Horndean Technology College	Hampshire	Community School	Secondary
Ixworth Free School	Suffolk	Free School	Secondary
JFS	Brent	Voluntary Aided School	Secondary
Ken Stimpson Community School	Peterborough	Community School	Secondary
Kettering Park Infant School	Northamptonshire	Academy Converter	Primary
King Edward VI Grammar School, Chelmsford	Essex	Academy Converter	Secondary
Lakenham Primary School	Norfolk	Foundation School	Primary
Lanchester Community Free School	Hertfordshire	Free School	Primary
Lansdowne School	Lambeth	Community Special School	Special
Linton Heights Junior School	Cambridgeshire	Academy Converter	Primary
Marfleet Primary School	Kingston upon Hull	Academy Converter	Primary
New Haw Community Junior School	Surrey	Academy Converter	Primary
Ormiston Park Academy	Thurrock	Academy Sponsor Led	Secondary
Our Lady and St Patrick's, Catholic Primary	Cumbria	Voluntary Aided School	Primary
Parkside Primary School	East Riding of Yorkshire	Community School	Primary
Parley First School	Dorset	Community School	Primary
Pennington CofE School	Cumbria	Voluntary Controlled School	Primary
Penryn College	Cornwall	Academy Converter	Secondary
Princetown Community Primary School	Devon	Foundation School	Primary
Ravensbourne School	Havering	Academy Special Converter	Special
Ringway Primary School	Northumberland	Community School	Primary
Round Diamond Primary School	Hertfordshire	Community School	Primary
Sir Robert Pattinson Academy	Lincolnshire	Academy Converter	Secondary

School name	Local authority	Type	Phase
Sir Thomas Boughey Academy	Staffordshire	Academy Converter	Secondary
Soar Valley College	Leicester	Community School	Secondary
Sompting Village Primary School	West Sussex	Community School	Primary
St Augustine's Catholic College	Wiltshire	Academy Converter	Secondary
St John the Divine Church of England Primary School	Lambeth	Voluntary Aided School	Primary
St Katherine's Church of England Primary School	Essex	Foundation School	Primary
St Mary Redcliffe and Temple School	Bristol	Voluntary Aided School	Secondary
The Barlow RC High School and Specialist Science College	Manchester	Voluntary Aided School	Secondary
The Bishop of Hereford's Bluecoat School	Herefordshire	Voluntary Aided School	Secondary
The Cottswold School	Gloucestershire	Academy Converter	Secondary
The Marston Thorold's Charity Church of England School	Lincolnshire	Voluntary Aided School	Primary
The North Halifax Grammar School	Calderdale	Academy Converter	Secondary
Trinity Catholic High School	Redbridge	Voluntary Aided School	Secondary
Upminster Infant School	Havering	Academy Converter	Primary
Westwood College	Staffordshire	Academy Converter	Secondary
Winklebury Junior School	Hampshire	Community School	Primary

** The list of schools visited was amended on 19 December as Harris Academy Beckenham was included in error. This has also led to a minor amendment in figure 25.