Department for Education

Evaluation of Subject Excellence Clusters: Final Report

Research Report

February 2017

Rebecca Allen, Meenakshi Parameshwaran, Dave Thompson, Mike Treadaway, Sam Sims – Education Datalab, FFT Education Ltd

Simon Burgess – CMPO, University of Bristol

GSR
GOVERNMENT SOCIAL RESEARCH
Social Science in Government

Contents

List of figures

List of tables

Executive summary

Introduction

In line with international best practice, data on school and subject department performance has become central to school accountability in England. It is used by Ofsted, governors and parents to judge schools and by headteachers and other senior leaders to judge the performance of departments and teachers. However, relatively little is known about how effectively middle leaders in schools, such as subject department heads, interpret performance information. This report summarises the findings of a study into the way that departmental heads use this information for the purposes of school improvement, including through conversation with other schools.

Aims of the project

The overall aims and objectives of the study, entitled the Evaluation of Subject Excellence Clusters (ESEC), are to evaluate whether heads of subject departments in secondary schools are able to use detailed information on their department to aid their decision-making. Through the study, heads of maths, English and science departments were given simple comparative information on examination entries and outcomes across benchmarked groups of similar schools. The information is intended to help them make decisions on issues such as examination entry for groups of pupils, the allocation of pupils and teachers to classes, curriculum and teaching decisions for sub-groups of pupils and the performance management of teachers within their department.

Research methodology

This research study required the participation of heads of departments from secondary schools in England. Figure 1 shows an outline of the overall research design. Prior to recruitment of heads of department to the study, every secondary school in England was allocated to one of 497 clusters of six reasonably similar and geographically proximate schools. 118 schools, each from a different cluster, were recruited to the trial. These recruited schools are referred to as '**actively participating**' schools and the other five schools in their cluster are referred to as '**passively participating**' schools. Despite not representing a truly random sample of all schools, the characteristics of 'actively participating' schools are nevertheless broadly similar to secondary schools in England as a whole.

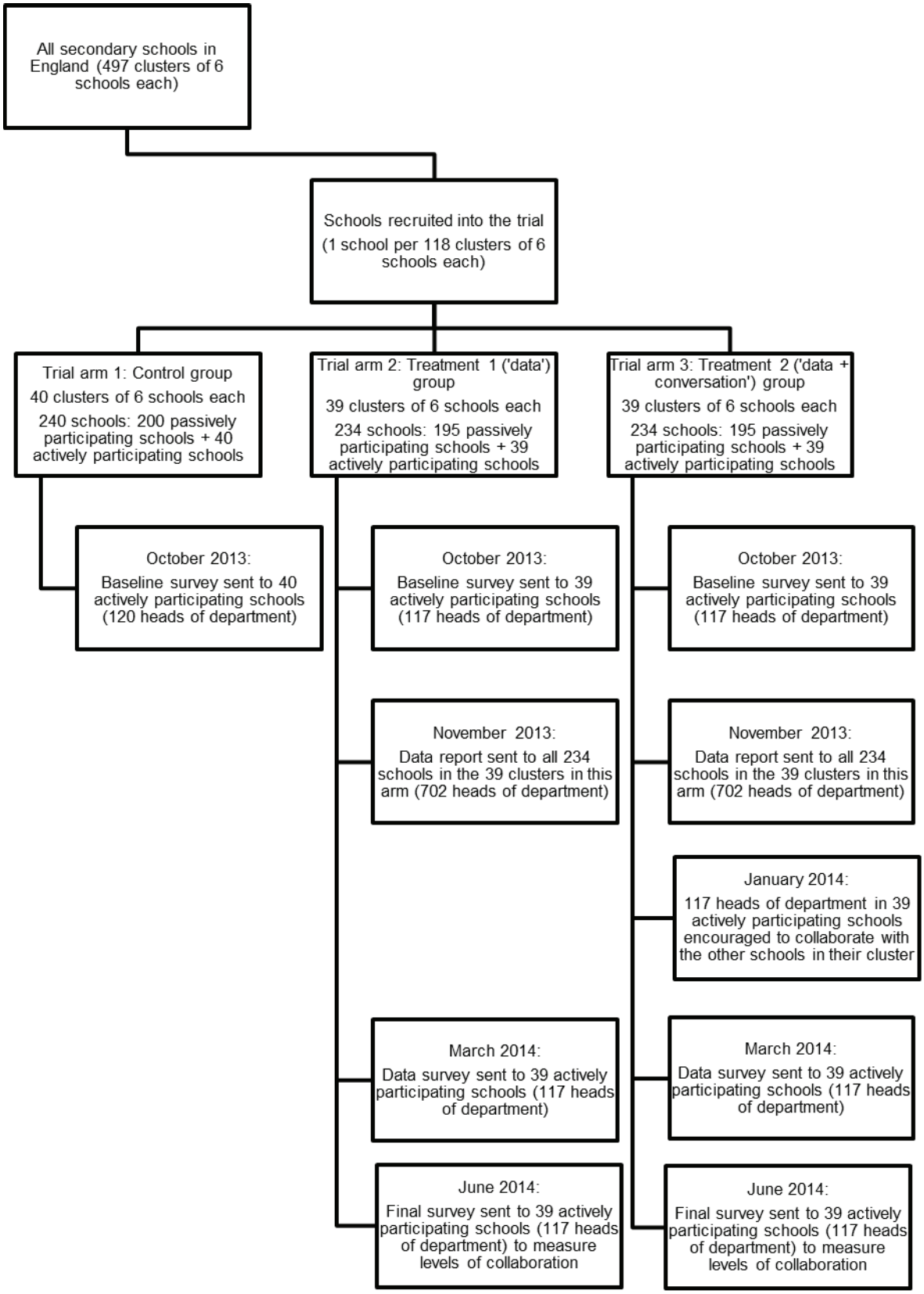
Each of the 118 school **clusters** were then **randomly allocated** into either 1) a **control** (n=40) group which received no intervention, or 2) a **treatment** (n=38) group which received information on their performance, relative to the other schools within their cluster, or 3) a separate **treatment** (n=38) group which received the same information but was also offered cash payments if they visited another school in their cluster and held a conversation with them. These three groups represent the **three arms** of the trial. As such, the research was designed as a **three arm cluster randomised controlled trial**, with the power to detect an effect size of one-third of a GCSE grade in each individual subject.

Heads of English, mathematics and science departments were recruited into the study from each of the 118 actively participating schools. This gave a total of 354 participants (120 in the control group, 114 in treatment group 1, and 114 in treatment group 2).

In October 2013, all 354 heads of department from the 118 actively participating schools were asked to complete a **baseline survey** to assess their pre-existing use and understanding of pupil performance data. Following this survey, all departmental heads in schools in the two treatment arms of the trial were sent a specially prepared information pack which allowed them to compare themselves to the five other named local schools in their cluster with similar demographic profiles. This approach was intended to provide easy to interpret information for staff, whilst maintaining fair and valid performance comparisons.

Two research surveys and performance measures from the National Pupil Database (NPD) were used to measure schools' behaviour before and after the intervention was made. The 118 actively participating schools from all three arms of the trial were invited to complete the baseline survey. Only the 76 actively participating schools from the two treatment arms of the trial were invited to complete the follow up **data survey** in March 2014. A final survey to summarise participation in the project was sent in June 2014. As well as measuring schools' behaviour before and after the intervention, the surveys also make it possible to assess the extent to which teachers engaged with the programme and gather feedback on the usefulness of the information provided.

Figure 1: Overall research design



Key findings

Effects of the two treatment arms

- Cross-sectional estimates of average treatment effects on attainment in each subject for 2014 and 2015 do not provide any evidence of the impact of treatment on schools participating in the 'data' treatment arm. This holds true if we compare 'actively participating' treatment schools to 'actively participating' control schools, or if we make a similar comparison using the 'passively participating' schools (Section 5.1).There was a low level of take-up of the 'conversation' part of the treatment by departmental heads in the 'data + conversation' treatment arm (treatment group 2). Only 11 schools had taken part in at least one conversation by the time of the final survey. This meant that it was not possible to estimate the effect of the 'data + conversation' treatment arm. Instead, results for the 'data + conversation' treatment arm were combined with those for the 'data' treatment arm.
- Three main reasons for the low level of take-up of the 'conversation' part of the treatment could be identified from survey responses and from email correspondence with the project manager. These were time constraints/ workload pressures, a lack of clarity about the purpose of the project and being unable to make contact with counterparts in other schools (Section 4.4).
- A statistically significant difference was found between entry rates into triple sciences in 'treatment' versus 'control schools' in 2015. However, this effect was driven by a fall in entries in the 'control' group, rather than an increase in entries in the 'treatment' group. The differences between entry rates for two sciences and English literature were not statistically significant (Section 5.3).

Target setting

- Schools generally first set GCSE targets for pupils either in year 7 (35% of departments), when they enter the school, or in year 9 (30%) and 10 (32%), when GCSE courses begin. More effective departments (departments with higher performance at GCSE) tend to set targets earlier, as do more effective schools (schools with higher performance at GCSE) (Section 3.1).
- In the majority of departments (57%), targets are set by senior leaders, with the next most significant single group being teachers (14%) and then specialist members of staff (10%) (Section 3.1).

Assessing pupil progress and departmental performance

- In both key stage 3 and key stage 4, just under 70% of departments reported assessing pupil progress at least once every half term. Effective departments assess pupil progress less often during key stage 3 but just as often during key stage 4 (Section 3.1).
- It was more common for pupil progress to be assessed at least every half term in maths (76% schools) and English (73%) than in science (62%) at KS3. At KS4 it was more common for pupil progress to be assessed at least every half term in English (77%), followed by science (70%), and least common for maths (63%) (Section 3.1).
- Almost all heads of department (96%) reported that they could confidently name the most effective teachers in their department. When asked how important pupil attainment data is in making these judgements about teacher effectiveness, 58% of respondents cited it as one important factor amongst others, and only 6% said it was the most important factor, with observation and appraisals also playing a part (Section 3.1).
- When asked about their own department's performance in the 2012/13 and 2013/14 GCSE cohorts, 43% of departments believed they performed better than their counterparts in schools with similar intakes, 31% believed they performed similarly, and just 16% thought they performed worse. The remaining 10% did not know how their department performed relative to their counterparts. This imbalance suggests that many departmental heads experience illusory superiority, the psychological phenomenon that people tend to overestimate their ability relative to others (Section 3.2).
- English departments have slightly more accurate self-perceptions of their performance relative to other departments in their schools, compared with maths and science departments (Section 3.2).

Differences between high and low performing departments

- Respondents from departments with better GCSE results (high performing departments) were more likely to rate themselves as superior, and less likely to rate themselves as inferior. Nevertheless, there was still a significant proportion of high performing departments who mistakenly believed themselves to be inferior (8%) and a remarkably high proportion of low performing schools who mistakenly believed themselves to be superior (29%), with another 43% believing they perform similarly to others (Section 3.2).
- Heads of department reporting that they 'didn't know' whether they were better or worse than equivalent departments in similar schools tended to be high performing, suggesting they may be less complacent. A similar result was found

when looking at heads of department comparisons with other departments in the same school (Section 3.2).

Interpreting performance data

- Heads of department were asked nine questions about the performance of their pupils within a particular prior attainment band. For each of these nine questions, 75% of departmental heads answered correctly as to whether their school performed statistically better or worse compared to schools like theirs (Section 3.3).
- Although 43% of departmental heads were able to correctly interpret performance information for all nine levels of pupil prior attainment, over 35% of departmental heads were only able to interpret six or fewer data points correctly. This shows that the majority (two thirds) of departmental heads are interpreting the majority (at least two thirds) of performance data correctly, but a substantial minority are not able to consistently interpret this information with accuracy (Section 3.3).
- Departments with more complex entry arrangements, i.e. science and maths departments who are considering additional qualifications for very high attainers particularly valued seeing general information on school exam entry profiles as provided in this study (Section 3.4).

Knowledge of and collaboration with neighbouring schools

- Departmental heads were asked about the characteristics and performance of 10 of their geographic neighbour schools. Just over 12% correctly identified all 10 of their geographic neighbour schools in terms of similarity of intake, with 66% correctly identifying at least 8. Around 11% correctly identified all 10 of their geographic neighbour schools in terms of whether or not it was high performing, with 71% correctly identifying at least 8. (Section 4.1).
- Although 89% of respondents had had professional contact with teachers from other schools (not necessarily cluster schools) over the past 12 months, only 39% reported that somebody in their departments had worked with a teacher from another school in order to improve the department (Section 4.2).
- Open text responses to the survey showed that departmental heads do explicitly seek out new collaboration opportunities with schools. Where departmental heads reported seeking out new collaborators, they formed the network through existing Academy chains and federations, with the help of exam boards, by contacting all local schools, by seeking out departments known to have strong results, or through personal contacts. Many report running subject networks with feeder primary schools (Section 4.2).
- Sixty-eight per cent of departmental heads agreed or strongly agreed that collaboration increased teachers' engagement with their work. Over 50% of

departmental heads also agreed or strongly agreed that collaborative activities increased teachers' confidence at work and their pedagogical skills. However, departmental heads were more sceptical about the effect of collaboration on teachers' morale at work (36% agreed or strongly agreed) and on their department's lesson resources (42% agreed or strongly agreed) (Section 4.3).

Conclusions

The research shows that heads of department are confident about their ability to interpret and use pupil performance data. They are also very familiar with performance data, including information comparing their department with others in the same school, and with equivalent departments in other schools. However, departmental heads tend to overestimate the ability and performance of their department relative to others. This tendency to overestimate performance could be kept in check by encouraging heads of department to collectively review each other's exam performance.

Providing schools with performance data and school exam entry information in alternative formats is useful to schools, particularly when it is clear and simple. However, providing information to schools is not sufficient to improve attainment outcomes or change entry patterns. Whilst schools are able to understand the information given to them, using it to aid decision-making and to implement change is more difficult.

The extent and nature of school collaboration in the England suggests that it is difficult to foster. Schools struggle to collaborate because of time constraints and workload pressures, a lack of clarity about the purpose of collaboration projects and being unable to make contact with counterparts in other schools. Additionally, heads of departments aren't always confident that collaboration has a positive impact on their teachers and their department. However, they are keen to know about the exam board choices and entry profiles of similar schools, and high performing departments also tended to be encouraged to develop relationships with other schools by senior leadership teams. This indicates that there is an appetite for knowledge sharing and that senior leadership teams influence the degree of collaboration taking place.

These findings should be of use in developing and refining policies which try to bring about school improvement through joint working arrangements, such as Teaching School networks and Specialist Leaders of Education. They can also help policy makers decide what types of information to provide to schools and in what format.

1. Introduction

Data on school and subject department performance has become central to school accountability in England. It is used by Ofsted, governors and parents to judge schools and by headteachers and other senior leaders to judge the performance of departments and teachers. However, relatively little is known about how effectively middle leaders in schools, such as subject department heads, can interpret performance information. This report summarises the findings of a study into the way that departmental heads use this information for the purposes of school improvement, including through conversation with other schools.

1.1. Aims and objectives

The overall aims and objectives of the study, entitled the Evaluation of Subject Excellence Clusters (ESEC), are to evaluate whether heads of subject departments in secondary schools are able to use detailed information on their department to aid their decision-making on issues such as examination entry for groups of pupils, the allocation of pupils and teachers to classes, curriculum and teaching decisions for sub-groups of pupils and the performance management of teachers within their department.

There are a wide range of delivery methods and metrics used by the main commercial providers of school performance information (i.e. Fischer Family Trust (FFT), Centre for Evaluation and Monitoring (CEM), GL Assessment). However, there exists a tension between comprehensibility and reliability in creating metrics that reflect school and teacher effectiveness. On the one hand, subject departments' raw GCSE results are straightforward for governors, headteachers and other senior teachers to understand, but they reflect the cohort's prior attainment and social mix as much as they do teacher effectiveness. On the other hand, sophisticated predictive scores and value added measures can be unintelligible to the untrained eye and so are rather underused by schools to aid decision-making. However, simple comparisons of exam entries and outcomes can be made across schools if restricted to a group of schools with close to identical pupil background characteristics.

In this research project, information on curriculum offer and exam outcomes across benchmarked groups of similar schools (i.e. clusters or families of schools) were devised and delivered to English, mathematics and science departments in 456 schools within 76 clusters. By directly delivering simple comparative information in an accessible format to departmental heads, the aim is to influence decision-making regarding choice of examination entry for groups of pupils, the allocation of teachers to classes and the performance management of teachers within their department. An

invitation to participate in a conversation and collaborative activities within the cluster of schools was sent to a random sample of half these schools to find out whether this type of knowledge mobilisation across peers in similar schools can be effective.

1.2. Existing literature on teachers' abilities to interpret and use data

A core aim of this research project is to evaluate the ability of school subject department heads to use detailed data and information on their department to aid their decision-making.

Data literacy for teaching

Existing research suggests that teachers' effective data use, known as data literacy of teaching, is dependent both on their data analysis skills *and* on their ability to make appropriate pedagogical decisions based on their interpretation of the data (Coburn & Turner, 2011). The requirement of two sets of knowledge (data and teaching) for effective data literacy for teaching means that, unsurprisingly, most research finds a lack of teacher capacity on effective data use in schools (e.g. Means et al., 2009). Some research goes as far as to claim that the majority of teachers do not use data properly, if at all, basing their decisions on intuition and limited observations instead (Schildkamp & Kuiper, 2010).

Data use in schools

Data use increases in schools with effective data systems, which make it easier to collate and analyse data (Schildkamp & Kuiper, 2010). However, merely having access to data is not the same as using it effectively to improve teaching and learning (Schildkamp & Kuiper, 2010; Stecker, Fuchs, & Fuchs, 2005). Data needs to be reliable, valid, relevant, accurate, timely and appropriate for the school user; otherwise, data use will remain at low levels (Schildkamp & Kuiper, 2010). Giving teachers time to collaborate and making data use a team effort may also increase data use, alongside school leaders instilling a vision, norms, and goals for data use in schools (Schildkamp & Kuiper, 2010). Training and ongoing professional development for school staff can contribute to more and better data use, although in some situations, appointing a data expert to facilitate data makes it easier for schools to use data (Kerr et al., 2006). Finally, data use increases in schools when accountability systems require the production of evidence of teaching and learning effectiveness (Ingram et al., 2004).

Teachers' interpretation of data

The way that teachers search for information and interpret data is influenced by their pre-existing cognitive frameworks (Honig & Coburn, 2008), particularly with regards to how they value and rely on different forms of evidence (Coburn & Talbert, 2006; Light et al., 2005). It is not easy for teachers to reconcile evidence gathered through standardised tests and examinations with their prior knowledge and experience of teaching and their pupils (Brown & Rogers, 2014). Rather, for teachers to be able to effectively use data to inform their practice, they need the time and resources to collaborate, multiple data sources (not just assessment data) to identify pupils' needs, and the ability to adjust their teaching practices accordingly (for a summary, see Jimerson, 2015).

Middle leaders and data use

It is particularly important for heads of department to be able to interpret and use data in their decision-making processes because of the key positions they hold in school social networks (Farley-Ripple & Buttram, 2015). Whilst classroom teachers typically use classroom level data to inform instructional practice and reflect on their own performance, school leaders use school level data to inform school policy decisions and support teacher professional development (Schildkamp & Kuiper, 2010). Heads of department need to be able to understand and make use of data both at the classroom level and at the school level. They often have more teaching experience and higher pedagogical knowledge than their classroom colleagues, leading them to be identified as "experts" and sought out to provide guidance and leadership (Farley-Ripple & Buttram, 2015). With this position comes the responsibility for heads of department to model effective data use and be supportive and enthusiastic of data use so as to encourage staff to do the same (Sutherland, 2004). This makes a lack of data literacy for teaching a particular concern for heads of department, as they transmit significant amounts of information between colleagues and are expected to provide guidance, knowledge and leadership. Teachers in "expert" positions are not complacent, however, believing instead that they have a lot to learn about data literacy for teaching and only reporting moderate levels of comfort in data use (Farley-Ripple & Buttram, 2015).

1.3. Existing literature on knowledge mobilisation

This research project aims to better understand how to encourage knowledge mobilisation both within and across schools. All the schools in the treatment arms of this study received carefully designed information packs intended to give them an improved understanding of their own performance. Additionally, schools in treatment group 2 were encouraged to engage in 'conversations' with comparable nearby

schools so as to share experience, insights and knowledge. (The networks they develop are intended to become the conduits along which knowledge is transferred (Muijs et al., 2010).

Knowledge mobilisation between schools

Since the introduction of Education Action Zones in the late nineties, there have been many policy initiatives in the UK designed to mobilise knowledge in this way. These include: Beacon Schools, Excellence in Cities, Specialist Schools, Leading Edge Partnerships, Advanced Skills Teachers, the London Challenge (and other City Challenges), Specialist Leaders of Education, National Leaders of Education, Academy Chains and Teaching School networks (Bell et al., 2006; Sims, 2012). Similar reforms have also been implemented in the USA (see, for example: Wohlstetter et al., 2003; Wohlstetter et al 2013).

But despite policy makers' long-standing interest in using networks to mobilise knowledge, very little is known about networks in education. Leanna and Pill (2006) and Meier and O'Toole (2001) are some of the only controlled studies of the effect of external social capital on pupil attainment. Both studies find that increased external social capital is positively related to reading and maths scores and Leanne and Pill (2006) also show this relationship to be mediated by instructional quality. In both cases however, the research is limited by the cross-sectional design and the potential for omitted variable bias. As Chapman and Hadfield conclude in their review of the field, "despite their rising popularity, the literature pertaining to their purposes, design and functions remains limited. Furthermore, their impact on schools, teachers and pupils is even less clear" (2010, p309).

Knowledge mobilisation in other sectors

Outside education there has been more systematic study of knowledge mobilisation networks. In particular, researchers working in the fields of Management and Organisation Science have conducted many studies investigating whether and how network links can facilitate the flow of knowledge. In their meta-analysis of the literature, Wijjk et al. (2008) find that inter-organisational knowledge transfer is associated with increased organisational performance and greater innovation. They also find that the number and 'strength' of relationships (how long they have been established, and how often interactions occur) have a positive and statistically significant effect on the amount of knowledge transfer that occurs. The primary importance of these factors has also been emphasised in two other, more narrative reviews (Inkpen and Tsang, 2005; Phelps et al., 2012).

However, the management literature also provides a number of warnings about how difficult it is to establish useful network links. Kane et al. (2005) found that superior

ideas can be rejected if they are suggested by an ‘outsider’. This was later confirmed with a separate experimental study (Kane, 2010). Trust is also an important determinant of whether ideas are adopted (Wijjk et al., 2008) across organisations, but building trust for the purposes of knowledge sharing requires time and effort (Abrams et al., 2003). Because the type of knowledge that teachers are likely to share with each other is also not easy to demonstrate (it may take considerable time and effort to verify its truth) it will be harder to transfer (Kane, 2010; Phelpss, 2012; Wijjk et al., 2008). Another reason to be cautious about the potential for this sort of knowledge mobilisation effort is that teachers are already extremely busy. The 2013 Teacher’s Workload Diary Survey found that the average secondary classroom teacher works 55 hours a week (DfE, 2014) which may leave them with little time to travel to others schools to engage in conversations.

In summary, there is some suggestive evidence that teacher’s external links can improve school performance, and a good deal of evidence that organisations in general perform better as result of external links. However, there are also reasons to doubt whether teachers will have the time to develop network links and a number of reasons why these links might not lead to productive knowledge exchange. The next section sets out the design of this research.

2. Design

2.1. Randomised controlled trials

The research was designed as a **three arm cluster randomised controlled trial**. A **randomised controlled trial** involves the random allocation of participants into either 'control' (no intervention) or 'treatment' (intervention) states. With sufficient sample size, this randomisation guarantees that the schools in each of these groups are identical, at least in expectation. It follows from this that any differences in outcomes between these groups can confidently be attributed to the intervention they received, rather than pre-existing differences between the groups.

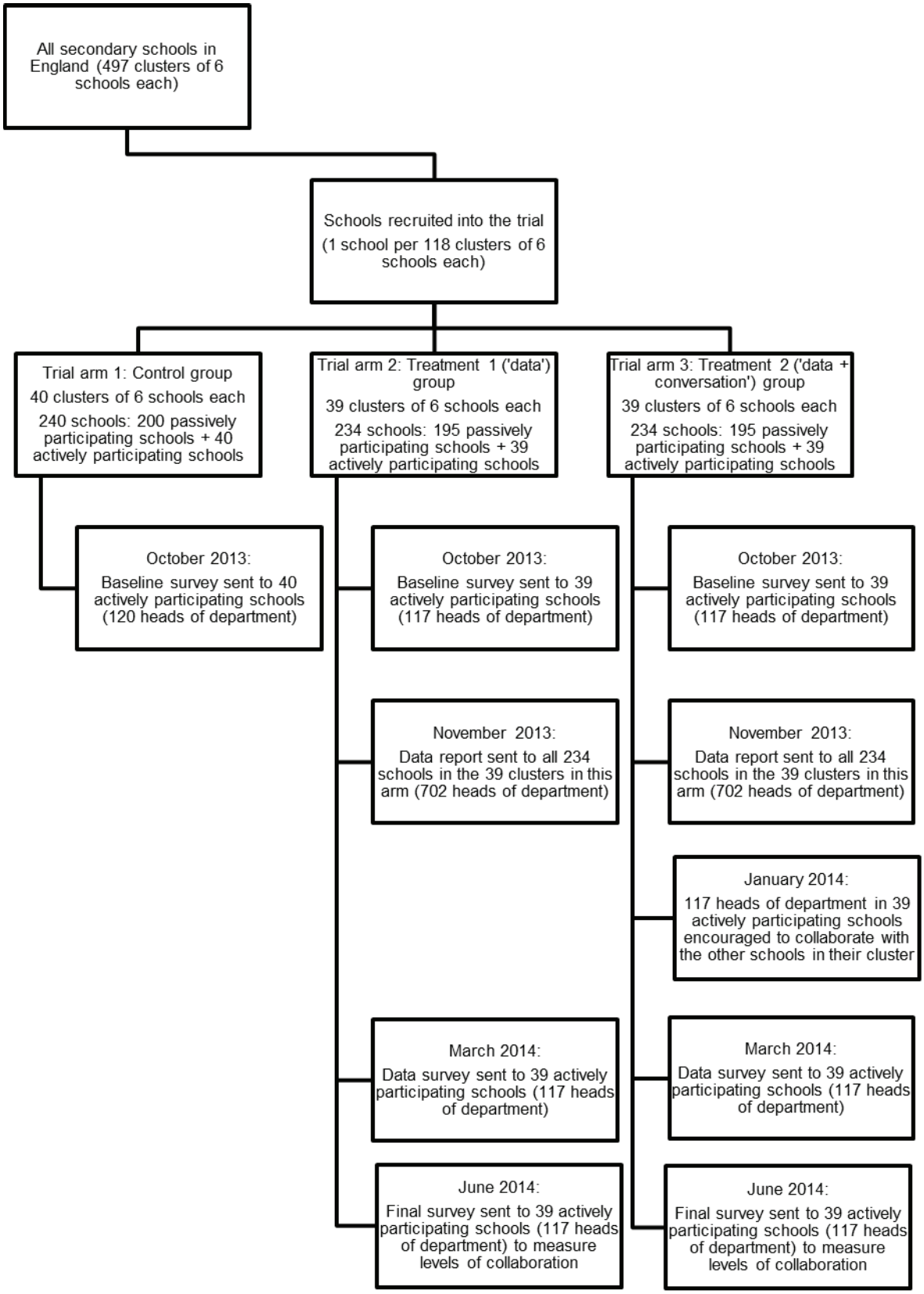
2.2. School clusters

Because the intervention being tested here involves schools collaborating with each other, the schools were allocated to treatment or control conditions as a group, or '**cluster**'. This is necessary because otherwise treatment group schools might end up unwittingly 'exposing' control group schools to the treatment by conducting a conversation with them. Prior to recruitment of participants, every secondary school in England was allocated to one of 497 clusters of six schools that are both reasonably close to each other and have similar demographic profiles (See Appendix 1 for methodology).

2.3. Three arms of the trial

Once schools were allocated to clusters, they were then placed into either 1) a control group which received no intervention 2) a treatment group which received information on their performance, relative to the other schools within their cluster, or 3) a separate treatment group which received the same information but was also offered cash payments if they visited another school in their cluster and held a conversation with them. These three groups represent the **three arms** of the trial. The overall research design can be seen in Figure 2.

Figure 2: Overall research design



2.4. School recruitment and treatment allocation

Recruitment and treatment allocation was run as follows. Of the 497 clusters initially created, 36 were deemed unusable due to the schools within the clusters being too distant to permit collaboration, or too dissimilar to permit comparison. From this remaining set of clusters, 118 schools were recruited, each from a different cluster, to take part in the trial. This was achieved through contacting schools, initially by letter and then by phone, and persuading them to take part. The 118 'actively participating' schools, along with the five 'passively participating' schools in their cluster, were then randomly allocated to one of the three arms of the trial. This was done on a close to even split, with 40 allocated to the control group and 39 to each of the information and conversation treatment arms.

2.5. Representativeness of the school sample

Despite not representing a random sample of all schools, Table 1 shows that the characteristics of those that volunteered to take part are nevertheless broadly similar to secondary schools in England as a whole. Three differences are visible:

1. there is a slightly lower proportion of pupils with high prior attainment (top quintile) among volunteer schools and a slightly higher proportion of pupils with low prior attainment (bottom quintile);
2. percentages of pupils eligible for free school meals (FSM) or with English as an Additional Language (EAL) are also slightly below average;
3. schools in Inner London and the North East are under-represented among the volunteers.

Table 1 also compares the characteristics of the clusters of schools in the control and treatment groups. Again, the overall balance appears good, however there are some differences visible including: a slightly lower percentage of low performing pupils in the control group; a slightly lower percentage of FSM eligible pupils in the information arm (Treatment_1) and some variation among the representation of ethnic groups and different regions.

Table 1 Characteristics of sampled schools

		Control	Treatment_1	Treatment_2	All Volunteers	All other schools	All schools
	Schools	40	39	39	**118**	2,862	2,980
	Number of pupils on roll (Jan 13)	21,773	22,952	22,892	**67617**	1,583,100	1,650,717
Prior attainment (Y9-Y11)	Mean fine grade	4.62	4.61	4.60	**4.61**	4.63	4.63
	% high prior attainment	18.39	18.66	17.98	**18.34**	19.79	19.73
	% low prior attainment	18.17	19.63	19.66	**19.17**	18.69	18.71
Pupil Context (Y9-Y11)	% FSM	14.54	13.45	14.39	**14.12**	15.38	15.33
	% EAL	11.98	11.03	11.59	**11.52**	13.33	13.25
	% SEN (SA+/Statement)	9.90	8.45	8.42	**8.91**	8.21	8.24
	% White British	79.16	80.72	79.39	**79.77**	76.45	76.59
	% Indian	1.26	1.21	1.18	**1.22**	1.36	1.36
	% Pakistani	2.39	2.24	1.83	**2.15**	2.40	2.39
	% Black African	2.19	1.43	2.03	**1.88**	2.94	2.90
	% Bangladeshi	3.38	4.12	2.94	**3.48**	3.37	3.37
Region	East of England	2.50	7.69	15.38	**8.47**	6.50	6.58
	East Midlands	17.50	15.38	10.26	**14.41**	12.68	12.75
	Inner London	5.00	0.00	0.00	**1.69**	4.30	4.19
	North East	5.00	0.00	2.56	**2.54**	5.07	4.97
	North West	17.50	15.38	15.38	**16.10**	14.29	14.36
	Outer London	7.50	7.69	15.38	**10.17**	9.57	9.60
	South East	10.00	20.51	15.38	**15.25**	15.79	15.77
	South West	15.00	10.26	10.26	**11.86**	9.78	9.87
	West Midlands	12.50	15.38	7.69	**11.86**	12.12	12.11
	Yorkshire & the Humber	7.50	7.69	7.69	**7.63**	9.89	9.80

Source: 118 volunteer schools and remaining 2,862 state maintained schools in scope (see Appendix 1)

The clusters of six schools are designed to share intake characteristics, but those involved in the treatment arms of the project will clearly be diverse in terms of attainment and this may affect their interpretation of the data sent to them. In order to understand the extent of the diversity of attainment within each cluster, the difference between overall average point scores (APS in English, mathematics and science was calculated at each school compared to an average of 30 other schools nationally that are similar to the cluster (referred to as the 30-similar-schools average). Table 2 shows the number of schools in each of the 78 treatment clusters where the difference between a school's overall APS is statistically significantly above or below the 30-similar-schools average. Statistically significant differences in APS between a school and the 30-similar-schools average for their cluster indicate that the differences in APS are less likely due to random chance and more likely due to real differences in the academic performance of pupils at the school.

In Table 2, clusters with less diversity of attainment are those clusters where most schools do not perform significantly differently to their 30-similar-school average. These will be the clusters found in the rows and columns for 0 schools and 1 school. In contrast, clusters found in the rows and columns for 3 schools or 4+ schools are those clusters where half or more of the six schools in the cluster perform significantly differently to their 30-similar-school average. These are the clusters with more diversity of attainment.

Table 2 shows that in science, in particular, there are some clusters with more diversity of attainment. In other words, science performance is not very similar within and across some clusters. For example, there are 15 clusters where at least four schools are significantly below average and none is significantly above average (the red cells). These are low performing clusters in science. Conversely, there are 8 clusters containing four schools that are significantly above average and where none is significantly below average (the orange cell). These are high performing clusters in science. In general, around half of the clusters in each subject (around 60% in science) contain at least one school with significantly above average performance and at least one school with significantly below average performance.

Table 2 Relative academic performance of treatment clusters compared to similar schools nationally

Subject	Number of schools in a cluster with an APS significantly below 30-similar-schools average	Number of schools in a cluster with an APS significantly above 30-similar-schools average					Total clusters
		0	1	2	3	4+	
English	0	1	7	8	5	1	22
	1	3	10	7	3	3	26
	2	5	10	3	1	0	19
	3	2	5	0	1	0	8
	4+	3	0	0	0	0	3
	Total	14	32	18	10	4	78
Maths	0	4	7	6	7	2	26
	1	8	6	7	3	2	26
	2	2	12	2	0	1	17
	3	2	3	0	0	0	5
	4+	2	1	1	0	0	4
	Total	18	29	16	10	5	78
Science	0	0	1	2	4	8	15
	1	0	4	9	7	1	21
	2	0	4	6	7	0	17
	3	4	5	1	0	0	10
	4+	11	3	1	0	0	15
	Total	15	17	19	18	9	78

Source: Volunteer treatment schools (*N* = 78 schools in 78 clusters) compared to similar schools nationally (*N* = 30 similar schools per cluster)

Data on the nature and extent of collaboration were collected through three surveys. In October 2013, a baseline survey was sent to all 354 departmental heads for English, maths and science in the 118 actively participating schools. This baseline survey measured the extent and nature of their collaborative activity prior to receiving the intervention. Departmental heads were offered the chance to win an iPad if they returned the survey by a certain date.

In the second stage of the research, the schools in the two treatment groups (the 'data' group, and the 'data + conversation' group) were sent the specially designed information packs (PDF data reports). The other five schools in their cluster were also sent the same information, along with an explanation of why they were receiving it. A second survey (data survey), sent in March 2014, then measured their understanding and interpretation

of this information. The data survey was sent to the 234 departmental heads at the 76 actively participating schools in the two treatment groups. Again, those who participated had the chance to win an iPad for themselves, or someone else in their department.

In the third stage of the research, the 39 actively participating schools in the 'data + conversation' treatment arm were offered a payment of £200 per head of department who took part in bilateral conversations with two of the five other schools in their cluster, and a further payment of £450 if they also took part in a tri-lateral conversation with three of the five schools in their cluster. A third survey (final survey) was then administered in June 2014 in order to measure any changes in the nature and extent of collaborative activity. The final survey was sent to the 234 departmental heads at the 76 actively participating schools in both of the treatment arms. However, responses to the third survey were so low that the findings could not be reliably used in this research report. This issue is discussed further in the following section.

2.6. Response rates of surveys

Three main surveys were administered during the project. The first survey, the baseline survey, was sent out to 354 heads of science, maths and English departments in the 118 participating schools. The intention of this survey was to measure the way in which schools use data on their performance and also the extent and nature of collaboration between schools prior to the intervention. The response rate for the survey was just under 64%, with comparable response rates for English, maths and science departments (further details on response rates can be seen in Appendix 2: Survey response rates). The full survey can be seen in Appendix 3: Surveys and data information sent to schools.

The second survey, the data survey, was sent to 234 heads of science, maths and English departments in the 78 treatment schools. The intention of this survey was to find out how well heads of department interpret data on the pupils they have recently taught. The research also sought to understand the extent to which heads of department know their neighbouring schools and have relationships with them. The data survey was administered shortly after sending departmental heads a four page PDF data report (in hard copy and by email). This four page PDF data report contained information about the six schools in the departmental heads' cluster, including information on exam entry patterns, exam outcomes, key stage 4 headline indicators, and school contextual characteristics. An example PDF data report can be seen in Appendix 3: Surveys and data information sent to schools. All the questions relate to the survey. The response rate for the survey was just over 42%, with better response rates from maths departments than from English and science departments. The full survey can be seen in Appendix 3: Surveys and data information sent to schools.

The third survey, the final survey, was sent to 234 heads of science, maths and English departments in the 76 actively participating schools in both treatment arms. The intention of this survey was measure any changes in the nature and extent of collaborative activity in the 'data + conversation' treatment group. Unfortunately, only 20 responses were received to this survey, giving a response rate of 9%. The responses provided indicated that only 11 departmental heads had taken part in at least one conversation, whilst eight departmental heads had not made contact with any of the other schools in their cluster. The very low response rate means that it is not possible to draw reliable conclusions on the impact of the 'data + conversation' from the final survey. Instead, results from both treatment arms had to be combined into a single treatment of 'data'.

2.7. Limitations of the design

The RCT required clusters of schools to be created with the aim of encouraging conversation within each cluster of similar schools. The approach was taken of actively recruiting just one school in each cluster, allowing all the others to act as 'passive' participants in the experiment. This had a major advantage that it made sign-up to the experiment possible; an alternative design was considered that would have required positive sign-up of multiple schools within each cluster but it wasn't considered likely that this would succeed at the sign-up stage.

The design chosen limited external validity in the sense that school sign-up was non-random, although as shown earlier, those participating schools are reasonably reflective of the overall population of schools. Internal validity of the experiment was high because full randomisation was achieved across the treatment and control arms.

On reflection, the advantages of creating this design with just one actively participating school in a cluster came at a major disadvantage in the operation of the 'data + conversation' treatment arm, where the single actively participating school found it difficult to persuade other schools in the cluster to take part in conversation activities. Low participation in the 'data + conversation' treatment arm meant that it was not possible to evaluate the impact of this treatment. Instead, results from both treatment arms had to be combined into a single treatment of 'data'. Further details on departmental heads' experiences and challenges of collaboration can be found in section 5.4.

3. Results – departmental use of performance data

The surveys were used to learn about how departmental heads use pupil performance data. The baseline survey asked about current departmental practices and also tested how well they know about their department's recent GCSE performance. The data survey tested whether they are able to correctly interpret the GCSE performance given to them.

3.1. Use of performance data by departmental heads

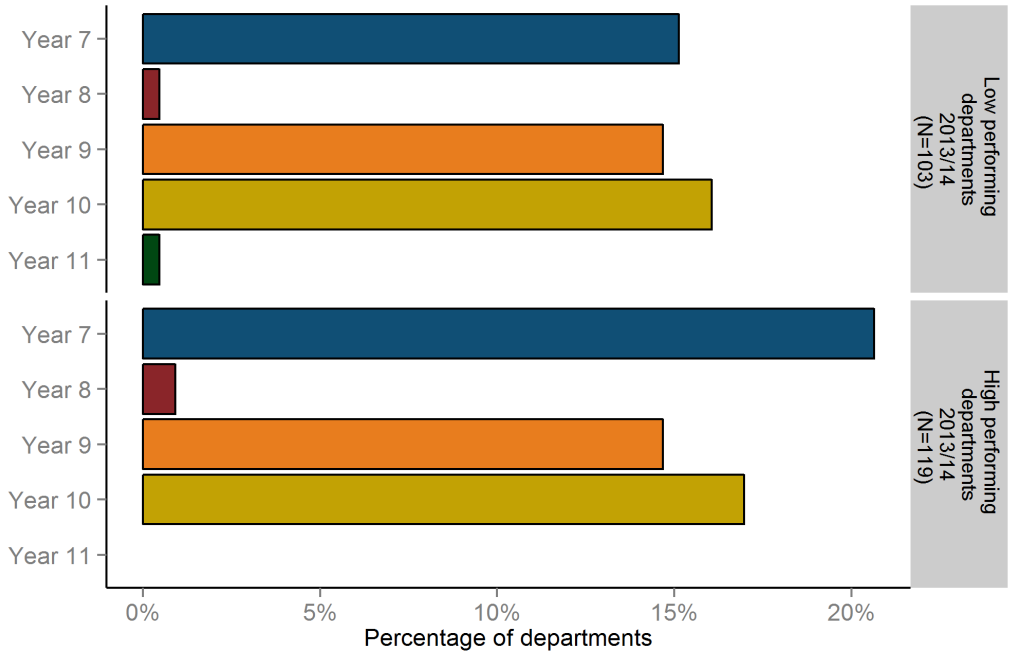
Pupil target setting

Departmental heads were asked about current practice in pupil target setting and assessment of progress. Schools generally first set GCSE targets for pupils either in year 7 (35% of departments), when they enter the school, or in year 9 (30%) and 10 (32%), when GCSE courses begin. Many who do not currently set explicit overall class targets report that they plan to do so in the future for performance management.

In order to explore this further, a new attainment measure (based on average point score at GCSE) was generated for each department in the sample controlling for a wide range of characteristics of their intake, including prior attainment, ethnicity, deprivation and special educational needs (i.e. a contextual value added measure for end of key stage 2 to end of key stage 4). The measure is not ideal, since it fails to control for other variables such as maternal education (Dearden et al., 2011) and migration patterns (Burgess, 2014), which have been shown to affect attainment but are not widely measured. In light of these limitations, schools and departments were split into just two categories, high and low performing. High performing schools (departments) are those with a positive contextual value added measure (CVA), whilst low performing schools (departments) are those with a negative CVA. In this analysis, high performing schools/departments are also known as more effective schools/departments, whilst low performing schools/departments are also known as less effective schools/departments.

Figure 3 shows the year in which targets are first set, by departmental effectiveness. It is evident that the more effective departments tend to set targets earlier, with noticeably more setting them straight away in year 7. None of the more effective departments wait until year 11 to first set targets. Splitting the responses by school effectiveness yields similar results.

Figure 3: Year in which targets are first set, by departmental effectiveness

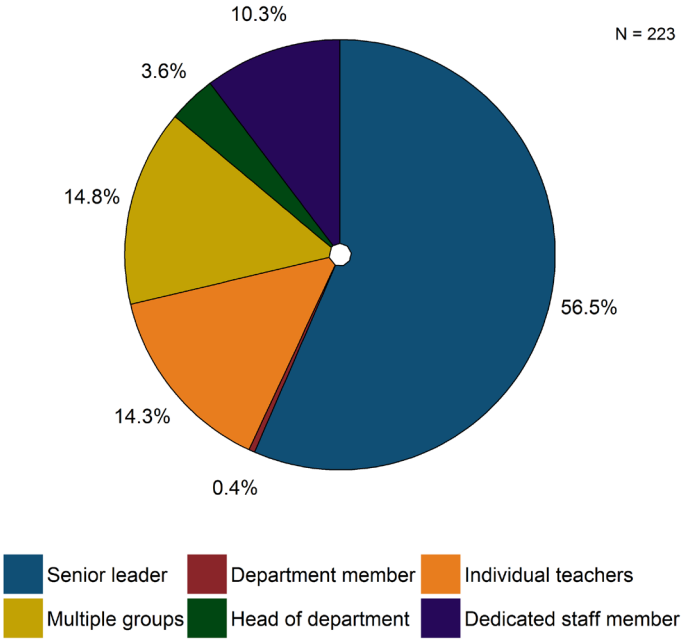


Source: Baseline survey (*N* = 222)

Departmental heads were asked about the systems and sources they use for target setting. The majority report using established assessment and prediction systems (e.g. FFT estimates, CEM predictions, GL assessment tests). Some manually apply DfE transition matrices to key stage results or use other in-house systems. Many schools report using a modification of the FFT estimates (e.g. "*FFT targets are moderated by the school*") or some combination of data. A few mention a data management system (e.g. Sisra or an internal system) but cannot report what the estimates entered into that system are based on.

Figure 4 shows that in the majority of departments (57%), targets are set by senior leaders, with the next most significant single group being teachers (14%) and then specialist members of staff (10%). The Head of Department sets GCSE targets in just 4% of departments. Responsibility for target setting either resides with the senior management team of the school or with the department themselves. Most schools report some objective basis for each pupil target, e.g. meeting FFT estimates, and many allow the final agreed target to deviate within reason. The process of agreement of the target occasionally involves the pupil themselves, but most frequently includes some opportunity for the class teacher to voice their opinion on the pupil's capabilities.

Figure 4: Who has primary responsibility for setting pupil targets?



Source = Baseline survey (*N* = 223)

Assessing pupil progress

Table 3 shows the frequency of departmental assessment of pupil progress. In both key stages 3 and key stage 4, just under 70% of departments reported assessing pupil progress at least once every half term. During key stage 3, more effective departments assess pupil progress less often than do less effective departments. However, during key stage 4, there are no differences in the frequency of assessment of pupil progress. One plausible interpretation of this is that more effective departments focus resources more on teaching up until year 10 or feel under less pressure to excessively monitor pupils.

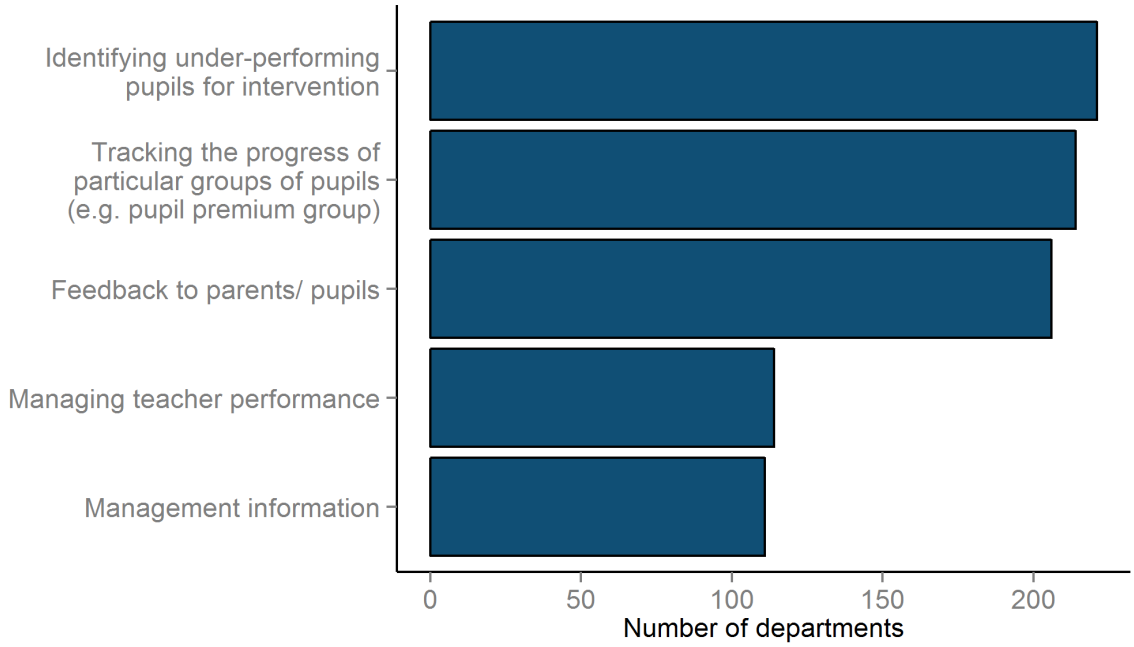
Table 3 How often do teachers assess pupil progress?

	KS3 Every half term or more	KS3 Every term	KS4 Every half term or more	KS4 Every term
English	73%	27%	77%	23%
Maths	76%	33%	63%	37%
Science	62%	38%	70%	30%
More effective departments	61%	39%	70%	30%
Less effective departments	74%	26%	69%	31%
All departments	67%	33%	70%	30%

Source: Baseline survey (*N* = 223)

A range of questions were asked to understand how this information was being used: respondents were asked to select all the options which applied. Figure 5 shows that the most commonly cited uses for pupil progress data were pupil-centred: identifying underperforming pupils, tracking progress of different groups, and providing feedback to pupils. Teacher-centred uses such as managing teacher performance or more general management information were cited around half as often. Beyond these categories, also mentioned were use of monitoring data for setting and grouping pupils, for allowing pupils to check their own progress against targets and for identifying the set of pupils who will contribute to meeting the floor standard.

Figure 5: What are any assessments of progress that you routinely collect used for?



Source: Baseline survey (*N* = 226 departments in schools)

To probe this further, respondents were asked whether they could confidently name the most effective teachers in their departments. The survey did not explicitly define what "most effective" meant. However, almost all of them (96%) said that they could. Respondents were then asked how important pupil attainment data is in making these judgements about teacher effectiveness and, although 58% of respondents cited it as one important factor amongst others, only 6% said it was the most important factor, with observation and appraisals also playing a part. Splitting the responses by departmental or school effectiveness reveals no obvious differences in the way that progress data is used.

Departments' plans to review current practices can be seen in Table 4. For all categories, only a minority of departments have no plans to review at all. Interestingly, six out of 10 respondents report that they are reviewing their policies for early entry at GCSE this year

(2013/14). This suggests that changes to Performance Tables with respect to first entries (rather than best entries) is encouraging departments to change their approaches to early entry. A similar proportion (between 60 and 70%) of departmental heads report that they are reviewing all categories this academic year (2013/14), except choice of examination board, where only 26% report any plans to review. This suggests that reviewing or changing examination boards is not an annual activity for most departments.

Table 4 Departments' review plans

	Reviewing this year (2013/14)	Reviewing in 2014/15	Reviewing beyond 2014/15	No plans to review
Exam boards specifications at KS4	26%	28%	7%	39%
Early entry for GCSE	60%	5%	2%	33%
Year 9 curriculum	70%	15%	2%	13%
Hours of teaching time in our subject	62%	10%	0%	28%
Development goals of our department	69%	13%	2%	16%
Allocation of teachers to classes	61%	5%	2%	33%
Substantial changes in our schemes of work	61%	21%	3%	15%

Source: Data survey (*N* = 60)

3.2. Interpretation of own performance

Departmental heads' understanding of their own performance was investigated by asking them to recall last summer's GCSE results. It is noteworthy that they were asked to do this at a time in Autumn term when the standard data management tools might not yet have key stage 4 data available. This might explain their relatively un-nuanced view of their own performance. For example, there was widespread mention of departmental outcomes in relation to last year's results and compared to national pass rates. Many departmental heads mentioned using exam board systems to evaluate results – this is obviously immediately available but doesn't allow comparisons across exam boards.

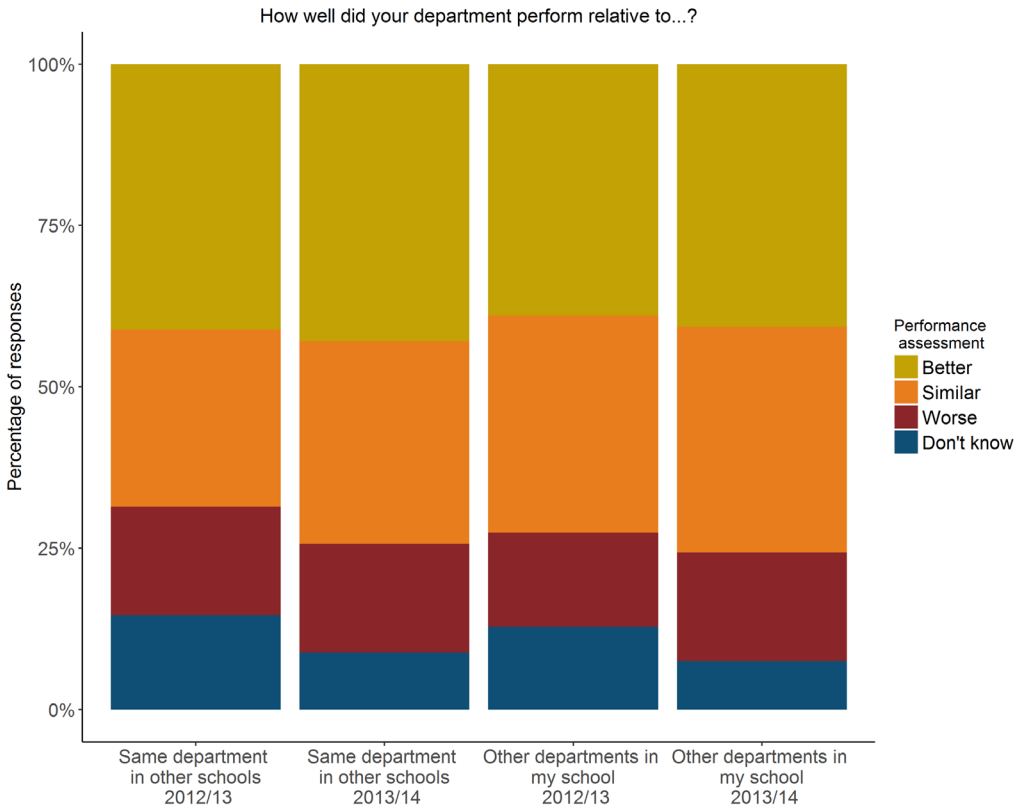
They were asked about how their own department performed in the past two years' GCSE cohorts relative to the same department in schools serving similar intakes. Given

that Table 1 suggests that the schools in the sample are broadly reflective of those in the country as a whole, it is reasonable to expect a similar proportion of schools to answer 'better than others', and 'worse than others'. However, as can be seen in Figure 6, the answers were actually skewed towards the more positive responses.

Comparison between schools

In 2013/14, 43% of departments believe they performed better than their counterparts in schools with similar intakes, just under a third (31%) believed they performed similarly, and just 16% thought they performed worse (Figure 6). The proportions were broadly similar for 2012/13 results, except for a larger proportion of 'don't knows', perhaps reflecting the number of respondents who only moved to their current school this year (see Figure 29). When asked to compare their department to others in the same school the results were similar, with a higher proportion than might be expected (41% in 2013/14) evaluating themselves as superior.

Figure 6: How well did your department perform relative to...

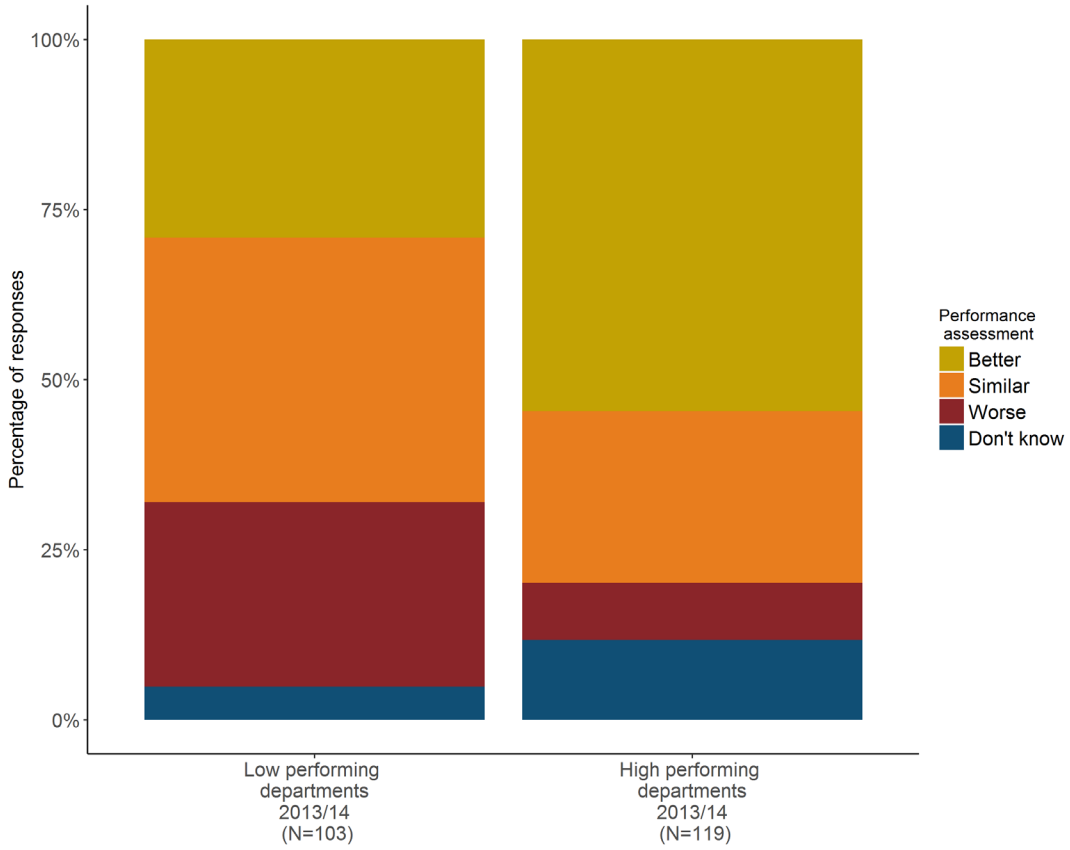


Source = Baseline survey (*N* = 226)

In Figure 7, it can be seen that those from high performing departments were indeed more likely to rate themselves as superior, and less likely to rate themselves as inferior. Nevertheless, there was still a significant proportion of high performing departments who mistakenly believed themselves to be inferior (8%) and a remarkably high proportion of

low performing schools who mistakenly believed themselves to be superior (29%), with another 43% believing they perform similarly to others.

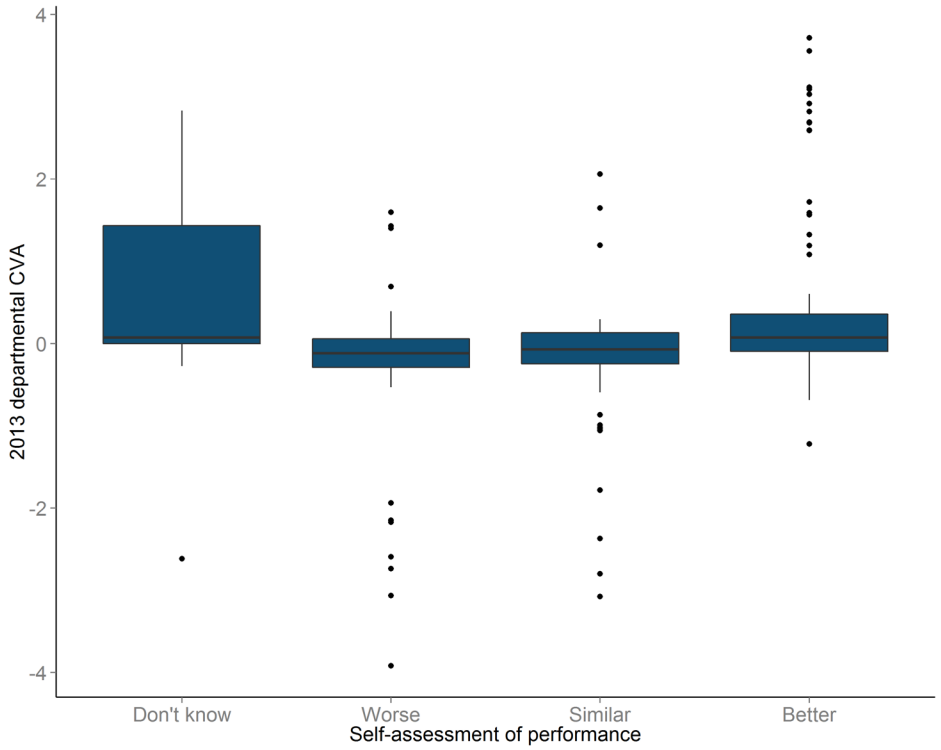
Figure 7: How well did your department perform relative to the same department in other schools?



Source: Baseline survey (*N* = 222)

In order to show the extent to which schools rated themselves differently to similarly performing schools, box plots were created for each judgment (Figure 8). The lowest point below each box indicates the 1st percentile in terms of Contextual Value Added (CVA) for each category, the lower end of the box represents the 25th percentile, the middle line in the box represents the median school, top of the box is the 75th percentile, and the uppermost point is the 100th percentile. Outliers have been excluded. A low performing school (for a given intake) will have a CVA below 0 and high performing school will have a CVA above 0. As can be seen from the substantial overlap between the 'worse', 'similar' and 'better' boxes, schools which perform similarly have widely varying perceptions of their own performance. Nevertheless, around three quarters of the departments describing themselves as either 'better' or 'worse' are accurate in their assessment. Interestingly, respondents that said they didn't know tended to be highly performing, suggesting they may be less complacent.

Figure 8: Distribution of departmental performance by self-assessment of performance relative to departments from other schools

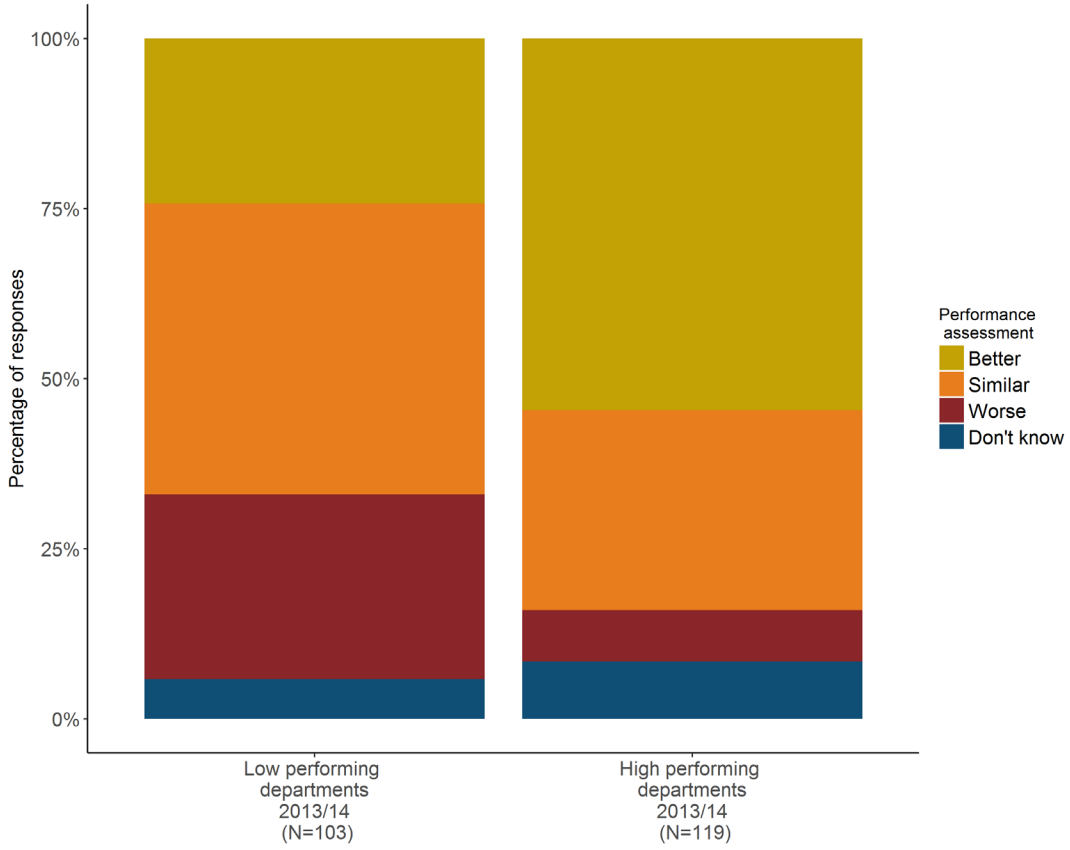


Source: Baseline survey (*N* = 222)

Comparison within schools

Departmental heads were then asked to make a second comparison with other departments in the *same* school, rather than the same department across other schools. When it came to comparing their department to the school's overall performance, respondents had slightly more accurate perceptions (Figure 9). More than half of high performing departments recognised themselves as such, and relatively few (8% in 2013/14) inaccurately believed themselves to be inferior. However, just under a quarter of low performing departments still believed themselves to be superior to others in their school. Low performing departments were also marginally more likely to recognise themselves as such. This is reflected in Figure 9, which shows that slightly more than three quarters of schools rating themselves as either 'better' or 'worse' are accurate in their assessment. Again, respondents that said they didn't know tended to be highly performing, suggesting they may be less complacent.

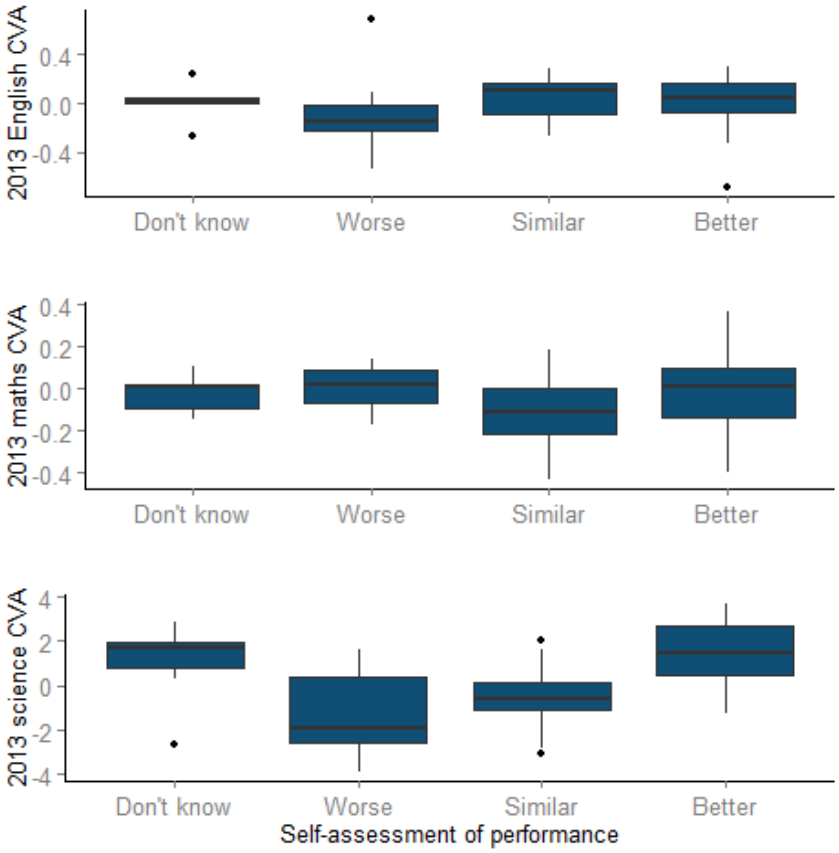
Figure 9: How well did your department perform relative to others in your school?



Source: Baseline survey (*N* = 222)

The responses are split by whether they were from English, maths or science departments. As Figure 10 shows English departments have slightly more accurate self-perceptions than maths or science, since only among English departments are more than 75% of 'better' and 'worse' judgements accurate.

Figure 10: Subject distributions of performance by self-assessed performance relative to other departments in same school

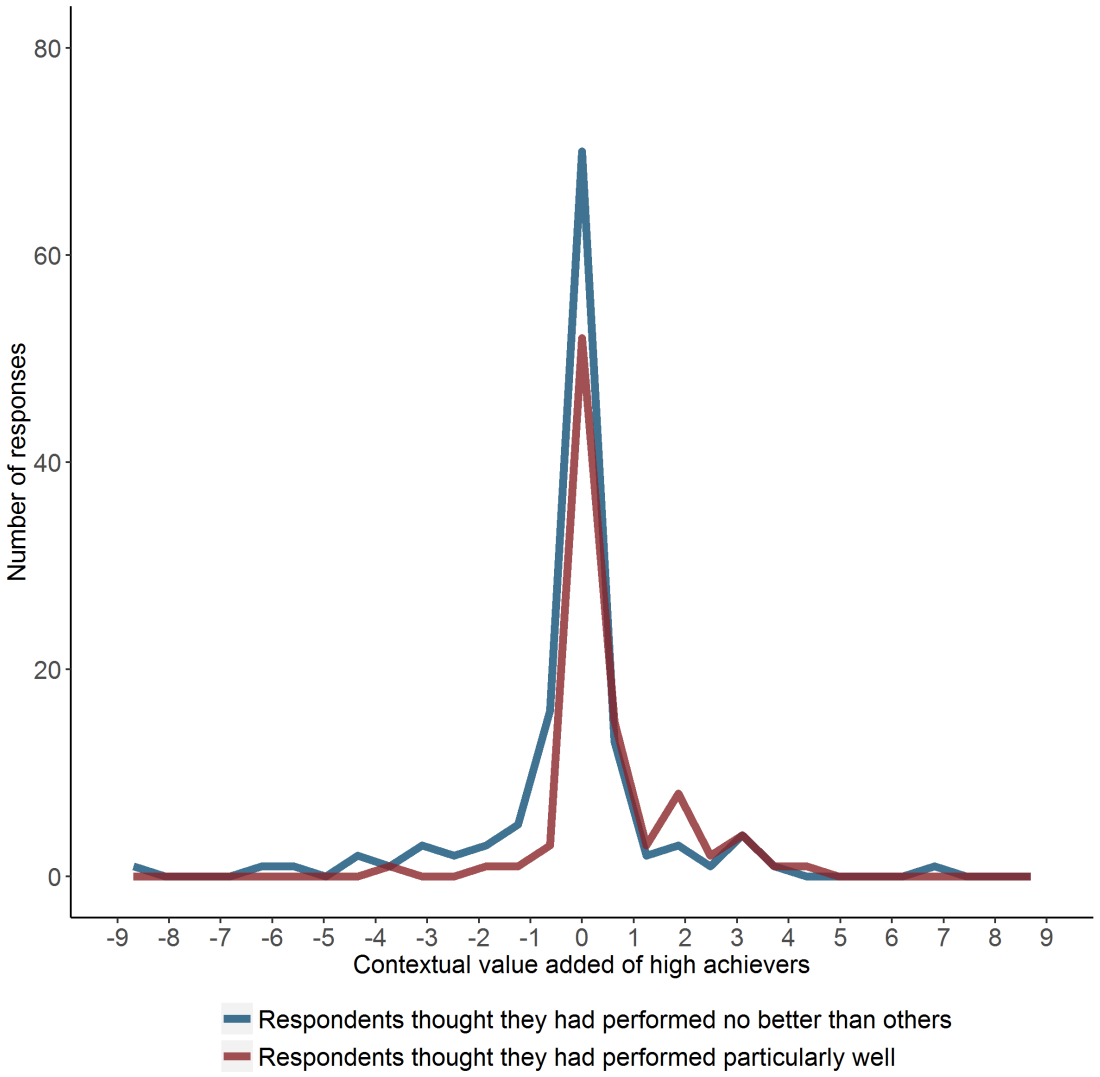


Source: Baseline survey (*N* = 222)

Comparison by ability group

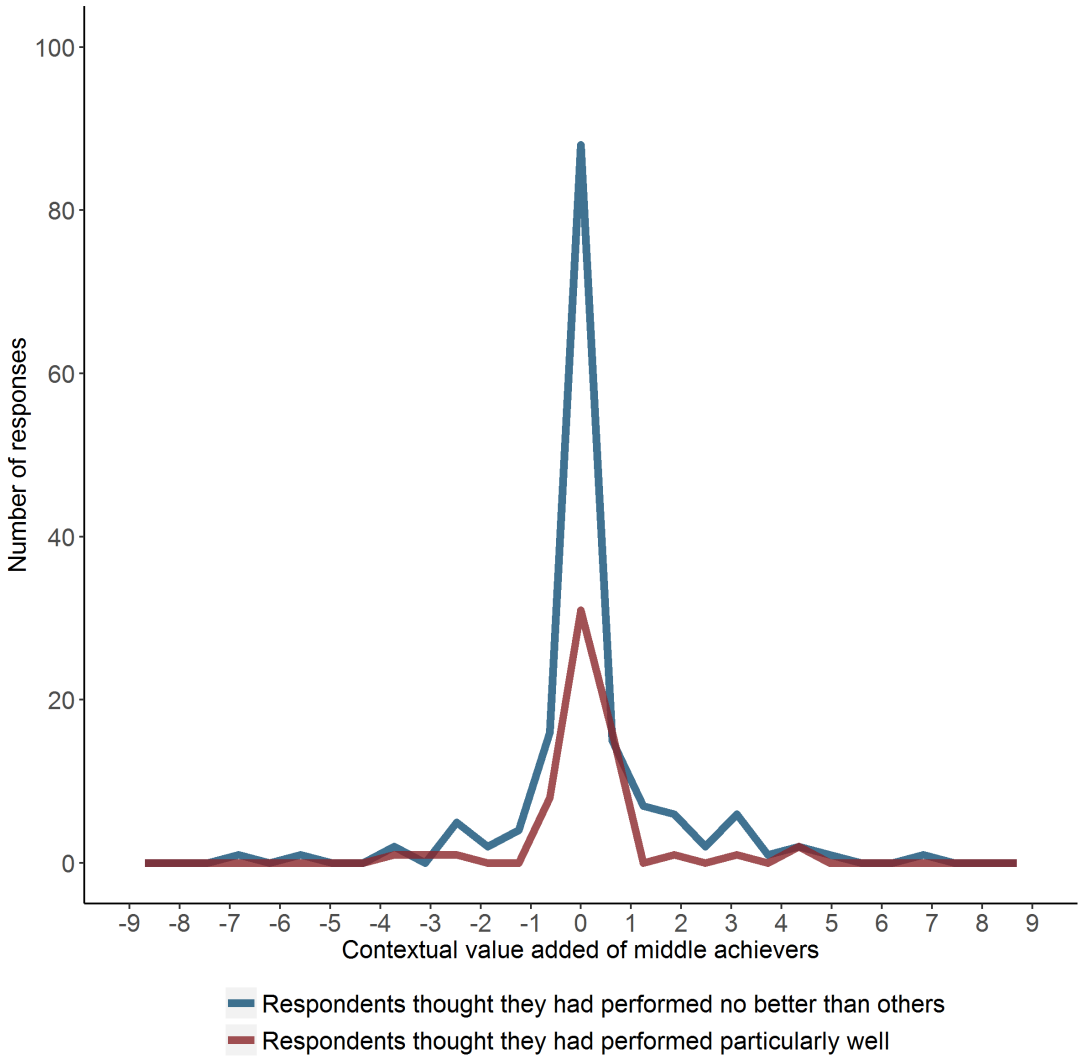
Finally, respondents were asked whether they thought any of their ability groups (high, medium, low) had done particularly well in the most recent set of GCSE results. The histograms in Figure 11, Figure 12 and Figure 13 show the distributions of the actual contextual value added for pupils in departments where respondents thought that group had done particularly well (red) against those who didn't (blue). The results suggest that respondents were better at identifying when their high and low performing pupils had done particularly well than they were for middle ability groups.

Figure 11: Contextual value added for high ability pupils where…



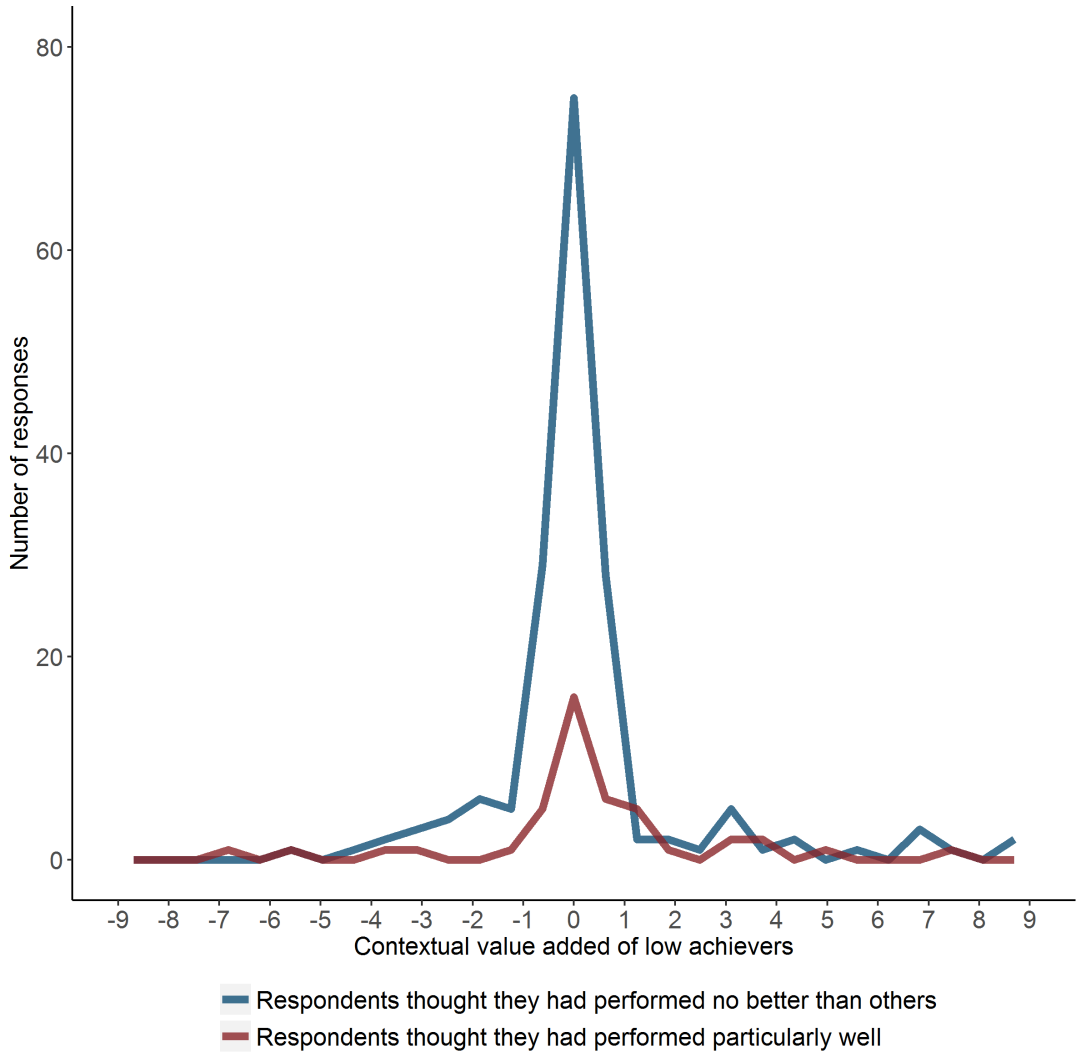
Source: Baseline survey (*N* = 222)

Figure 12: Contextual value added for middle ability pupils where…



Source: Baseline survey (*N* = 222)

Figure 13: Contextual value added for low ability pupils where…



Source: Baseline survey (*N* = 222)

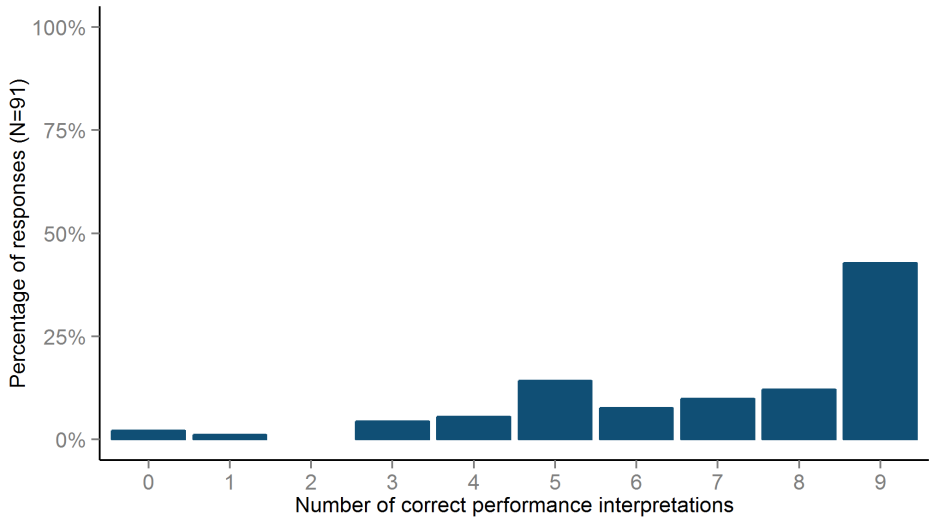
Overall, according to the data collected in this research, many departmental heads experience illusory superiority, the psychological phenomenon that people tend to overestimate their ability relative to others (Dunning et al., 2004). The dominant reason for this phenomenon is the desire to avoid undesirable judgements (Brown, 2012). This urge is particularly strong when assessment is of important tasks, such as teaching. The data collected in this research suggests that some middle leaders are simply choosing to ignore the data. However, recent research offers pointers on how illusions of superiority can be kept in check (Dunning et al., 2014). Middle leaders could help maintain objectivity by collectively reviewing each other's exam performance, such as through an annual inter-school meeting to analyse results. The Families of Schools database, which groups schools with similar intakes, can help make these comparisons more transparent. But what really matters is that performance is reviewed by respected, knowledgeable colleagues.

3.3. Proficiency in interpreting performance data

As part of the data survey given to departmental heads, some information on their recent GCSE performance was displayed in a relatively unfamiliar manner. The survey asked them questions about the interpretation of their performance data. For each of the nine questions about performance for pupils with a particular prior attainment band, about three quarters of departmental heads answered correctly as to whether their school performance was statistically better or worse compared to schools like theirs. The proportion of departmental heads that correctly interpreted the performance of pupils with different levels of prior attainment did not seem to vary across the different levels of prior attainment.

However, Figure 14 shows that departmental heads are not able to consistently interpret performance data. Although 43% of departmental heads were able to correctly interpret performance information for every level of pupil prior attainment, over 35% of departmental heads were only able to interpret six or fewer data points correctly. This is unlikely to be due to questionnaire fatigue as this was the first question in this survey.

Figure 14: Distribution of total correct interpretations of pupil performance data



Source: Data survey (*N* = 91)

Heads of department were asked: *"Look at the performance of your school's group of pupils with level 4A prior attainment at KS2. Which one of these statements best describes their performance at GCSE?"* Seventy per cent of respondents were able to correctly interpret the performance of their pupils with prior attainment of level 4A. Of the 30% who interpreted the information incorrectly, over half perceived their level 4A pupils' performance as being at least half a grade better than the national average.

3.4. Feedback on presentation of data

Almost all departmental heads found the information presented to be clear and many reported that they like the presentation of the data. Where difficulties in interpretation were reported, they related to not understanding the relationship between average point score and GCSE grades.

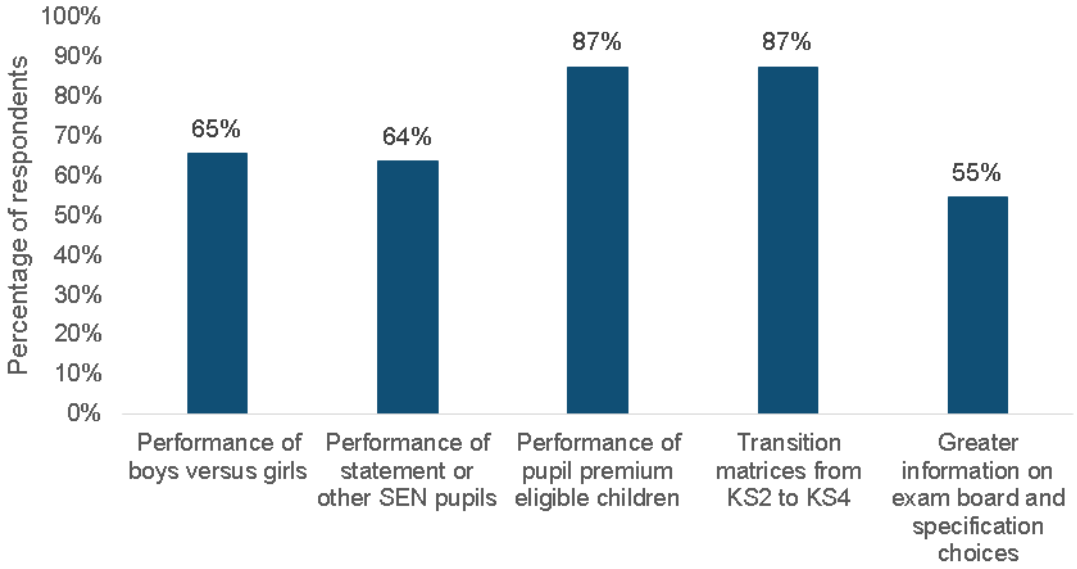
Most heads of department were familiar with their department's results and did not feel that this alternative presentation of the data told them anything new about the performance of their department the previous summer. However, a significant minority expressed surprise at how well or how badly a particular sub-group of pupils performed and many found it useful to compare their results to similar schools.

When asked which other details about the entry choices of similar local schools departmental heads would find useful, only a minority of heads said they would not be interested in learning more about the entry profile of similar schools. Most departmental heads would like to see information about other schools' choice of examination boards. When asked whether they would like to see any other information on exam entries, 76% wanted to see information on which exam specifications other schools have chosen to use, and 64% wanted to see how exam entry profile varies across different types of pupils (e.g. Special Educational Needs (SEN) or FSM pupils) in similar schools.

Departmental heads also overwhelmingly reported that they valued seeing general information on school exam entry profiles. This seemed to be particularly interesting for departments with more complex entry arrangements, i.e. many science departments and maths departments who are considering additional qualifications for very high attainers. The questionnaire was administered at a time of huge change – with reductions in the numbers of early entrants and removal of BTECs from performance tables.

Departmental heads also reported that they would find data on the performance of pupil premium children and data showing transition matrices from KS2 to KS4 particularly useful (Figure 15).

Figure 15: Other data on department or school performance that departmental heads would find helpful



Source: Data survey (*N* = 55)

The final page of the report displayed data on the relative performance of core departments within their school. Most reported that this information was familiar to them, reinforcing the evidence that within-school across-department comparisons of performance are consistently made. However, a few reported surprise that departments with a good reputation internally did not have particularly strong exam performance.

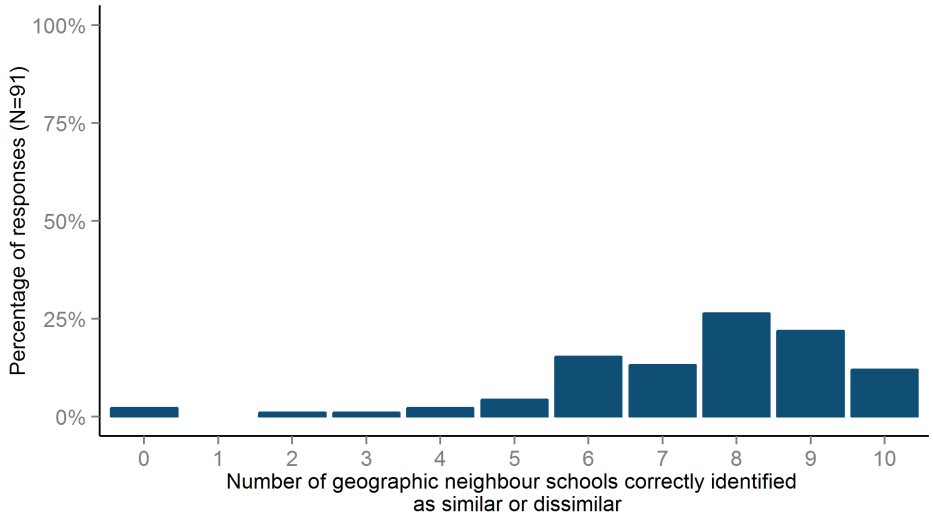
4. Results – knowledge mobilisation between schools

This section summarises the attitudes and current practice of departmental heads around collaboration or conversations with those working in other schools. Analysis is based on the two surveys, which asked questions to test their knowledge of other local schools and learn about any recent contact they may have had with these schools.

4.1. Knowledge of neighbouring and cluster schools

Departmental heads were asked about the characteristics and performance of 10 of their geographic neighbour schools. Figure 16 shows that, of 91 departmental head responses, just over 12% correctly identified all 10 of their geographic neighbour schools in terms of similarity of intake (defined as %FSM and key stage 2 Average Point Score (KS2APS) being within 0.5 of a school-level standard deviation). Around two thirds of departmental heads correctly identified the similarity of intake of at least eight of their geographic neighbour schools. It is important to note that departmental heads were not given a definition of "similar intake" and so survey responses deemed to be incorrect may be due to different interpretations of the phrase "similar intake".

Figure 16: Distribution of number of geographic neighbourhood schools correctly identified as having a similar intake

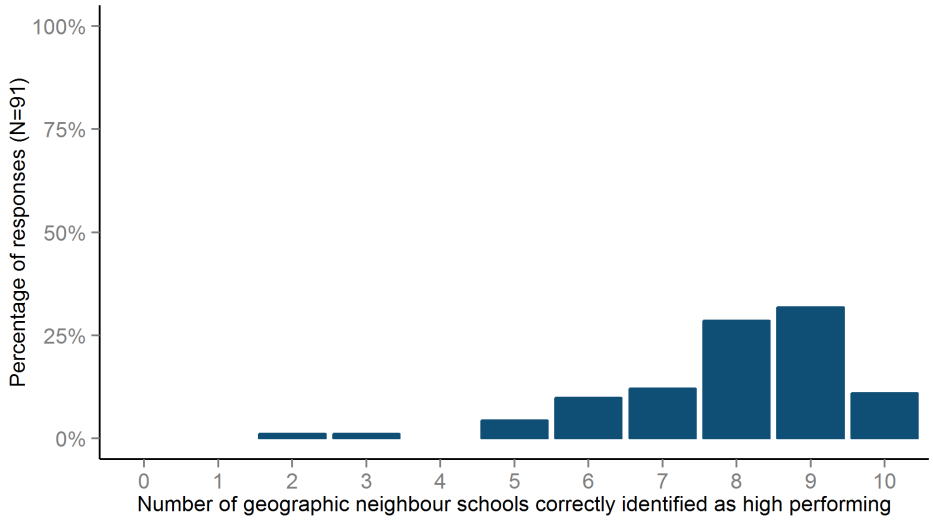


Source: Data survey (*N* = 91)

It can be seen in Figure 17 that around 11% correctly identified all 10 of their geographic neighbour schools in terms of whether or not it was high performing (high performing schools are those in or above the 75th percentile of three year average GCSE performance 2011/12-2013/14, as measured by the percentage of pupils achieving at least 5 GCSEs at A*-C grade including English and maths). Just over 71% of

departmental heads correctly identified at least eight geographic neighbour schools in terms of whether or not they were high performing. It is important to note that departmental heads were not given a definition of "high performing" and so survey responses deemed to be incorrect may be due to different interpretations of the phrase "high performing". However, interpretation is restricted to school-level attainment measures rather than department-level because the question focuses on school performance rather than department performance.

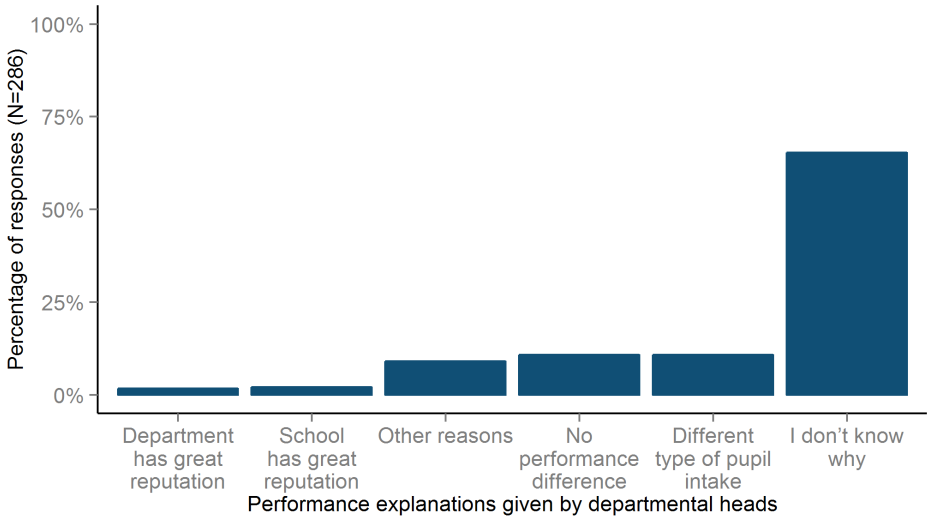
Figure 17: Distribution of number of geographic neighbourhood schools correctly identified as high performing



Source: Data survey (*N* = 91)

Departmental heads were asked if they could explain significant differences in performance in the 2013/14 GCSE performance tables between their department and their cluster schools' departments. Figure 18 shows that of the 286 responses received to this question, the majority (187) were "I don't know why". This indicates that most departmental heads responding to this survey cannot explain why the schools in their cluster performed as they did.

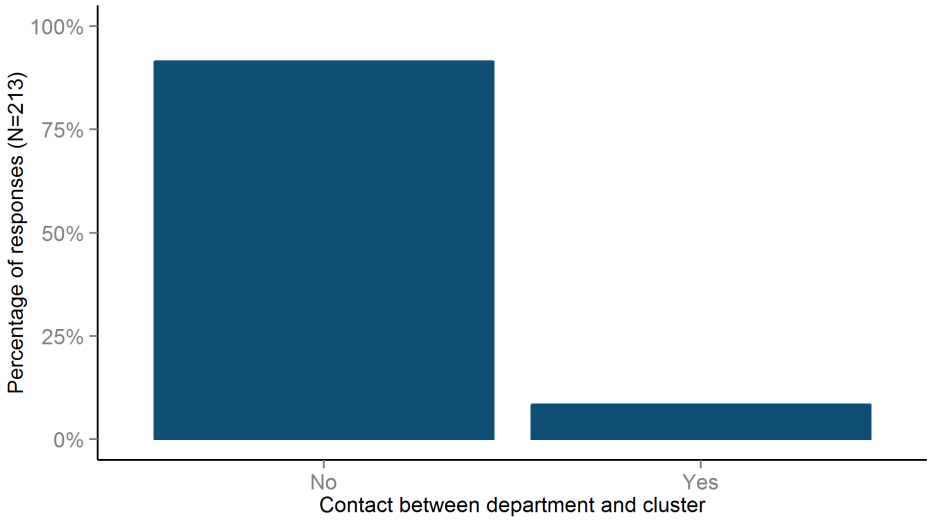
Figure 18: Departmental heads' performance explanations for geographic neighbourhood schools



Source: Data survey (*N* = 286 responses from 91 departmental heads)

Departmental heads were also asked if any previous contact had been made with the equivalent departments at their five cluster schools. Of the 213 responses received to this question, only 18 were positive (Figure 19). This suggests that most departments did not have contact with teachers in the equivalent department in their cluster schools.

Figure 19: Proportion of departments that have had contact with equivalent departments in their geographic neighbourhood schools



Source: Baseline survey (*N* = 213)

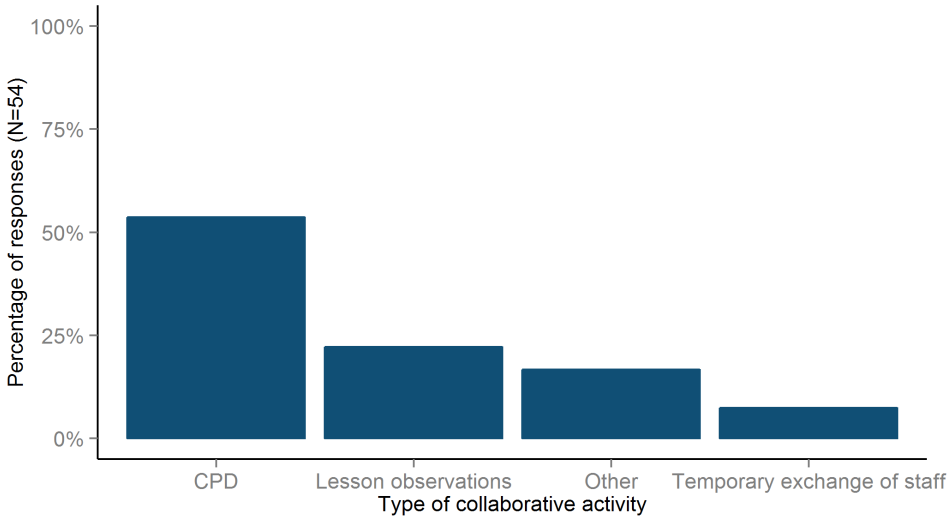
4.2. Existing contact with other schools

The baseline survey was also used to gauge the extent of collaboration in schools prior to any intervention occurring. The vast majority (89%) of respondents had themselves had professional contact with teachers from other schools over the past 12 months. In order to try and isolate the instances of contact which were designed to promote knowledge exchange, heads of department were then asked whether anybody in their departments had worked with a teacher from another school in order to improve the department. This reduced the number of positive responses to 39%, suggesting a significant amount of interaction with other schools was for purposes other than learning from each other. Schools were asked to describe the nature of their contact with other schools. Commonly mentioned reasons that could be interpreted as knowledge mobilisation and learning from each other included: general discussion (25 mentions), courses/CPD (25 mentions) and sharing good/best/professional practice (21 mentions). But there were also many mentions of meetings (96 mentions), many of which were for administrative purposes. Splitting the responses by how effective the department is revealed no clear differences.

The survey also asked whether they had ever explicitly sought out a new collaboration with schools. Where they had, they formed the network through existing Academy chains and federations, with the help of exam boards, by contacting all local schools, by seeking out departments known to have strong results, or through personal contacts. Many report running subject networks with feeder primary schools.

Figure 20 shows responses to the survey question *"which types of collaborative activities have you engaged in with other schools over the last three years?"* Respondents could select as many options as relevant. Of the 54 responses given to this question, the majority (29) related to collaborative CPD activities. The purpose of contact with other schools is most frequently only loosely defined e.g. observations of best practice or sharing of ideas. Where discussions are focused, departmental heads report discussing exam specification choices, coursework and moderation, curriculum development, and resource sharing. In general, the responding departmental heads do not report engaging in much collaborative activity with other schools at all over the last three years.

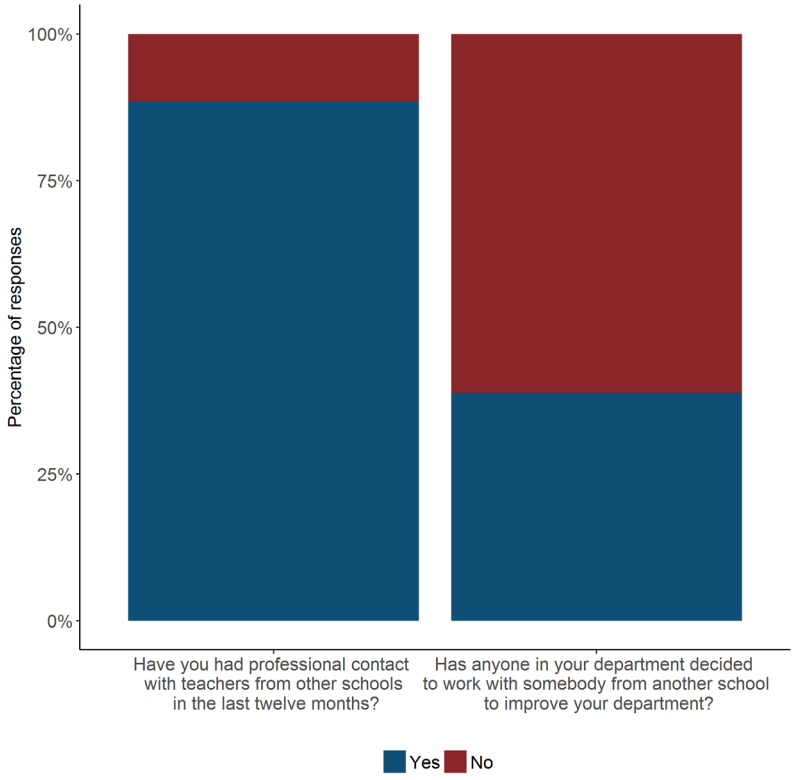
Figure 20: Types of collaborative activities engaged in over the last three years



Source: Baseline survey (*N* = 54)

Figure 21 shows that although collaborative activity was not frequently reported by heads of department, most respondents did report professional contact with teachers in other schools within the last 12 months.

Figure 21: Professional contact with teachers from other schools and collaboration with equivalent departments in other schools in the last twelve months

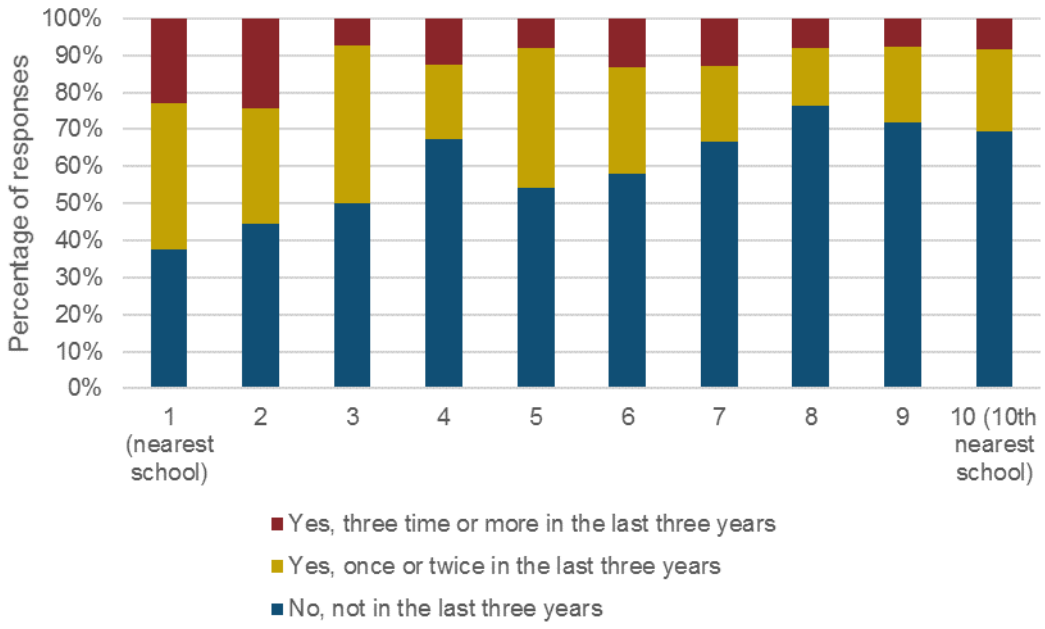


Source: Baseline survey (*N* = 219)

From the open text responses given to this question, a very wide range of types of contacts with teachers at other schools were reported. A large proportion of subject heads report that the local authority arranges regular opportunities for subject leads to meet. Others take part in similar events that are organised through subject networks, exam boards, federations, Teaching School Alliances or other groupings. Many report they are part of a smaller ad-hoc cluster of schools (e.g. 3-7 schools) who arrange meetings. Less frequently, schools are part of research projects or have arranged bilateral visits to observe lessons. A number of head of departments continue to work collaboratively with former colleagues from other job posts.

Figure 22 shows the distribution of responses given by departmental heads when they were asked if anyone in their department had collaborated with teachers in the equivalent department in their geographic neighbour schools. Responses are fairly positive, with between 16% and 43% of departmental heads reporting collaborating once or twice with a neighbouring school in the last three years.

Figure 22: During the last three years, has anyone in your department collaborated with teachers in the equivalent department in the following schools?



Source: Data survey (*N* = 402 responses from 91 schools)

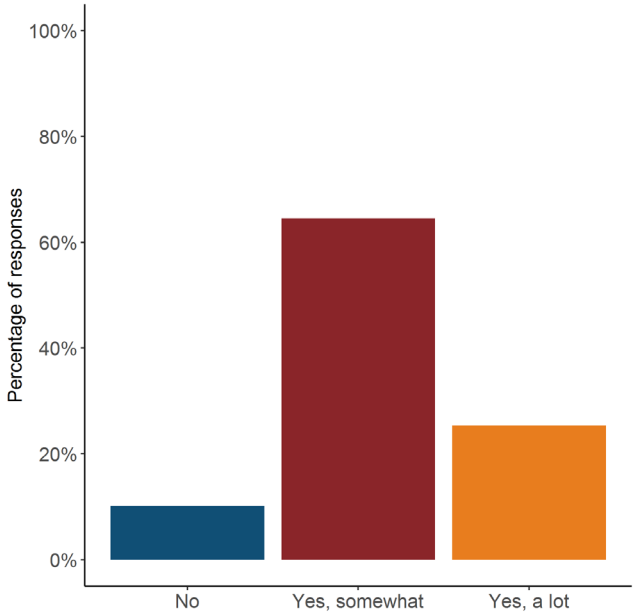
Departmental heads were asked about pre-existing contact with equivalent departments in their five cluster schools. The vast majority (over 80% of each of the five named cluster schools) reported no contact at all in the last three years. Though the clusters were designed to achieve geographical and intake similarity, it is noteworthy the infrequency of contact between schools beyond immediate proximal neighbours. Just five departmental heads reported contact with a cluster school three times or more over the last three years.

Though departmental heads had little contact with cluster schools, they did recognise these schools as sharing similar pupil intakes (about 70% agreed) and thus similar likely departmental interests. Despite this, few departmental heads said they might make contact with these cluster schools during the remainder of the academic year (about 1-in-10 said they would definitely and 1-in-20 said they might).

4.3. Management attitudes to school collaboration

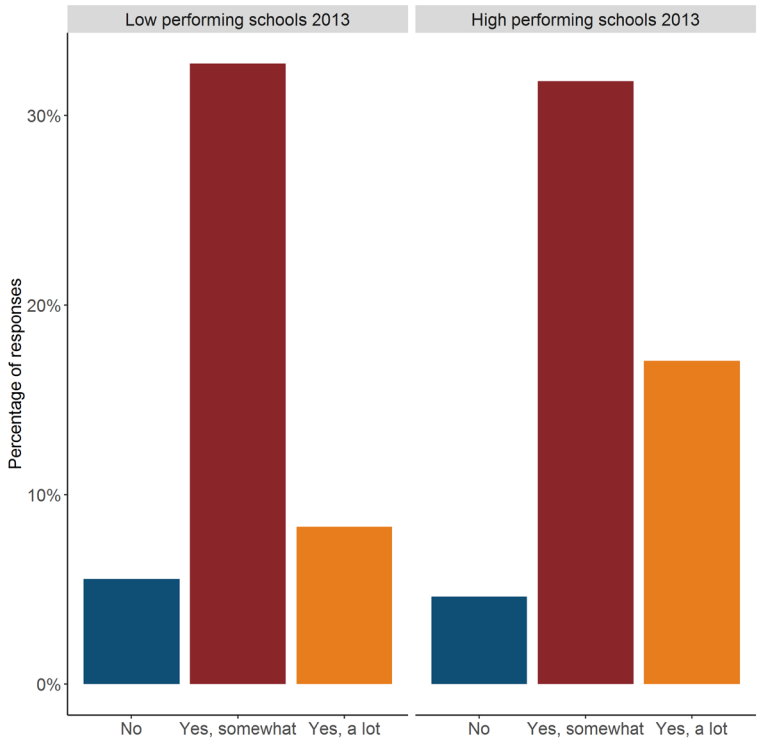
Heads of department were asked whether they were encouraged by their senior leadership team (SLT) to cultivate relationships with other schools. As Figure 23 shows, a quarter of respondents said they were encouraged to do so a lot, with 10% saying they were not. High performing schools were twice as likely to say they were encouraged a lot to cultivate collaborative relationships and were also less likely to receive no encouragement. However, splitting the data by school effectiveness (Figure 24) showed no difference in the proportion of departments who had actually been involved in at least some collaboration aimed at knowledge exchange. High performing schools were more likely to respond that SLT encouraged them to cultivate relationships with other schools or departments 'a lot'.

Figure 23: Does the SLT encourage you to cultivate relationships with other schools or departments?



Source: Baseline survey (*N* = 220)

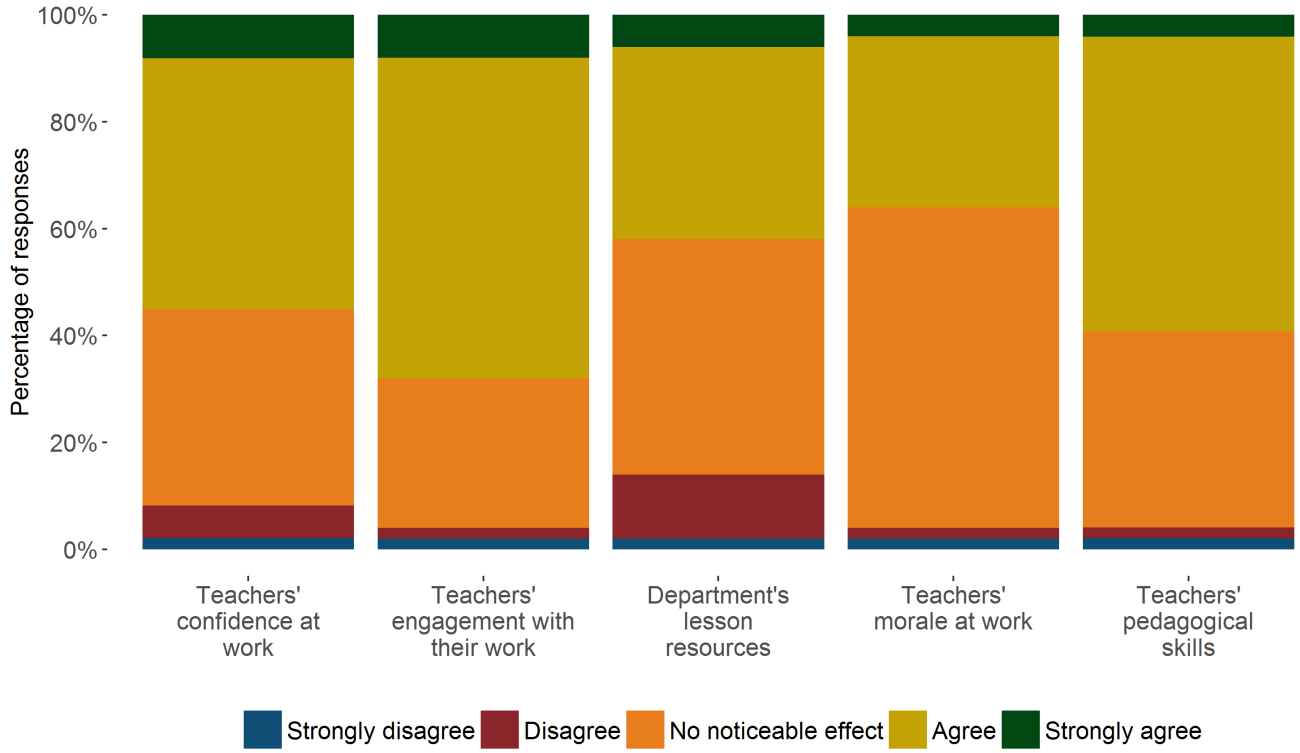
Figure 24: Does your SLT encourage school collaboration?



Source: Baseline survey (*N* = 101 low performing schools and 116 high performing schools)

Figure 25 shows the distribution of responses to a series of questions posed to departmental heads about the benefits they perceive of collaboration. Sixty-eight per cent of departmental heads agreed or strongly agreed that collaboration increased teachers' engagement with their work. Over 50% of departmental heads also agreed or strongly agreed that collaborative activities increased teachers' confidence at work and their pedagogical skills. However, departmental heads were more sceptical about the effect of collaboration on teachers' morale at work (36% agreed or strongly agreed) and on their department's lesson resources (42% agreed or strongly agreed).

Figure 25: Collaborating with other schools has increased our…?



Source: Data survey (*N* = 51)

4.4. Implementation of the conversation treatment arm

Departmental leaders in the 'data + conversation' treatment arm were invited to hold conversations with other schools in their cluster to support their performance and/ or continuing professional development. The conversations could be either 1-to-1 or in small groups consisting of at least three other schools. Brief guidance was provided about holding conversations and incentive payments offered of up to £650 per departmental leader per school. The project manager also offered to act as an intermediary in setting up conversations.

Incentive payments were to be made on the basis of a final survey in June 2014. This revealed a rather low take-up. Twenty responses (out of 114) were received from departmental leaders and found that 11 had taken part in at least one conversation. Eight of the departmental leads that responded to the final survey had not made contact with any of the other schools in their cluster. The remainder had attempted to contact an average of two schools each.

Three principal reasons for not contacting schools could be identified from survey responses and from email correspondence with the project manager. These were time

constraints/ workload pressures, a lack of clarity about the purpose of the project and being unable to make contact with counterparts in other schools.

Firstly, whilst some respondents affirmed a degree of support for the project, they had been unable to find the time to contact cluster schools due to workload pressures:

> *"I am still interested in this project and I apologise for not being more forthcoming. We had a disappointing Ofsted and a troublesome follow-up visit this year and everything at [the school] became very single-minded."*

Secondly, although the headteacher or deputy headteacher with curriculum responsibility had volunteered for the project, its aims were not necessarily shared by departmental leads:

> *"I am unsure what you would want me to say to the other schools."*

Finally and predominantly, schools struggled to make contact with other schools. This is not a new problem, of course (Sturgis et al., 2006). Without a direct contact for their opposite number in another school, departmental leads in participating schools often had to resort to leaving messages with the School Office and hoping for a response:

> *"I have emailed 3 different schools in the Subject Excellence Cluster and have asked for my details to be passed on to the Heads of English to arrange a meeting. Despite sending multiple requests, I have had no reply from any of the schools."*

> *"I really like the idea of it, the information is exactly what is needed to understand exactly how schools use entries in varying qualifications and for which pupils and with what degree of success. I really want [our maths + science] teams to run with this and [we] have been trying very hard to establish meaningful contact with those other schools in the cluster of 6, but this has been the project's downfall so far."*

Although participating schools had volunteered for the project and were incentivised to participate, the same was not the case for cluster schools. In many cases, participating schools were trying to make contact with schools with which no previous relationship existed. This may explain the lack of response.

5. Results – impact on GCSE results

The overarching aim of the evaluation is to see whether encouraging better use of data and conversations about performance within and between school departments ultimately leads to better pupil performance at GCSE. In this section estimates are presented of the impact of the treatment on GCSE attainment and entry patterns in English, mathematics and science using pupil data from the National Pupil Database between 2010 and 2015[1].

The three principal outcomes measured are:

- Points score in qualifications eligible for the English component of the English Baccalaureate
- Points score in qualifications eligible for the maths component of the English Baccalaureate
- Points score in qualifications eligible for the science component of the English Baccalaureate

Indicators of average point score in GCSEs in other subjects and entry rates in GCSE English literature, GCSEs in any two science subjects and GCSEs in the triple sciences (biology, chemistry, and physics) were additionally calculated.

There were major changes to the accountability framework for secondary schools during the period in which outcomes were observed (see The Wolf Report 2011). These changes included:

- restricting the set of qualifications that could be counted
- counting a pupil's first, rather than best, result in a subject result
- changing the GCSE equivalence of some non-GCSEs so that no qualification was counted as more than one GCSE
- limiting the number of non-GCSEs counted per pupil to two.

These changes had a major impact on the 2014 Performance Tables. Of all these changes, the second had the greatest impact, with average point scores in GCSE mathematics in state-funded mainstream schools falling between 2012/13 and 2013/14 following several years of increases (DfE, 2015).

Two approaches are taken to estimating treatment impacts: pupil cross-sections and school panels. In the pupil cross-sections, the attainment of pupils attending treated schools were compared with those attending control schools conditional on a set of pupil-level covariates, key stage 2 average point score, free school meal eligibility and gender. The basic model specification is shown in Figure 26. X_j represents the set of covariates

[1] 2014/15 data based on provisional results

for each pupil *p* in school *s*. μ_{ps} is the associated error term. Separate models are estimated for each year for each subject.

Figure 26: Pupil cross-sections model specification

$$y_{ps} = \delta treat_s + \sum_{j=0}^{J} \beta_j X_{jps} + \mu_{ps}$$

In the school panels, the change in outcomes within each school following the intervention were estimated. The basic model specification is shown in Figure 27. This involves creating a set of fixed school effects for each school (α_s) and a set of fixed year effects for each year (α_t) relative to the first year of the analytical window (2009/10). X_{jst} represents a set of time-varying covariates for each school (mean prior attainment, %FSM, %EAL). μ_{st} is the error term for each school *s* in year *t*. Finally, $treat_{st}$is a boolean flag set to 'true' for treatment schools in the year following the intervention (2012/13) and all subsequent years. The value of δ is therefore a difference-in-difference (DID) estimate of the impact of the treatment.

Figure 27: School panels model specification

$$y_{st} = \alpha_s + \alpha_t + \delta treat_{st} + \sum_{j=0}^{J} \beta_j X_{jst} + \mu_{st}$$

5.1. Average effect of treatment on attainment, all schools

Table 5 presents cross-sectional estimates of treatment effects in each subject for 2014 and 2015. All schools which were involved in the trial as active or passive participants are included. There is no evidence of the impact of treatment on schools participating in the 'data + conversation' treatment arm.

By contrast, a statistically significant effect is found for maths in 2015 in the 'data' arm. However, even though conventional statistical significance is achieved, the evidence of an effect is rather weak. Firstly, multiple comparisons are made in this report and so it is reasonable to expect a small proportion to achieve statistical significance simply by chance. Secondly, the difference of 0.5 points compared to control schools is equivalent to just 0.04 of a standard deviation. Finally, a Bayes Factor of 0.68[2] for the t-value shown indicates a relatively weak level of evidence against the null hypothesis that there is no

[2] Uses R package BayesFactor with default priors and a 'wide' scale for the prior distribution. There were 234 schools in the treatment group and 240 schools in the control group.

difference between treatment and control schools (Jeffreys, 1961). Broadly speaking, the alternative hypothesis is 1.5 times more likely.

Table 5 Cross-sectional effects of treatment on attainment by year and subject

Treatment Arm	Year	Subject	Diff	SE	t-value	p-value
Data	2014	Maths	0.37	0.21	1.75	0.08
	2014	English	0.22	0.25	0.89	0.37
	2014	Science	0.22	0.23	0.94	0.35
	2015	Maths	0.51	0.21	2.47	0.01
	2015	English	0.40	0.22	1.74	0.08
	2015	Science	0.37	0.22	1.71	0.09
Data + Conversation	2014	Maths	0.01	0.21	0.07	0.94
	2014	English	0.05	0.23	0.23	0.82
	2014	Science	0.04	0.23	0.16	0.87
	2015	Maths	0.12	0.20	0.60	0.55
	2015	English	0.07	0.24	0.30	0.76
	2015	Science	0.19	0.22	0.88	0.38

Source: Performance data for all schools in the trial (2,980 schools for each year)

5.2. Average effect of treatment on attainment, actively and passively participating schools

The random allocation of schools to treatment and control groups, together with additional controls for observed differences between schools, made it possible to make causal inferences about the effect of the intervention on outcomes.

The group of 118 volunteer schools are referred to as *actively participating schools*, and it is this group which provides internal validity for this research.

As relatively few of the school departments in the 'data + conversation' treatment arm took part in a conversation (13 out of a possible 114) both treatment arms are combined into a single treatment group and in Table 6 treated schools are contrasted with control schools among active participants. There is no effect of the treatment on outcomes.

Table 6 Cross-sectional effects of treatment on attainment, actively participating schools

Year	Subject	Diff	SE	t-value	p-value
2014	Maths	-0.25	0.43	-0.58	0.56
2014	English	0.03	0.53	0.06	0.95
2014	Science	-0.25	0.46	-0.54	0.59
2015	Maths	-0.21	0.49	-0.43	0.67
2015	English	0.06	0.40	0.16	0.87

2015	Science	-0.33	0.46	-0.72	0.47

Source: Performance data on actively participating schools (*N* = 118)

It may well be the case that those schools which volunteered for the project tended to have a positive predisposition towards performance data. However, the aim is to generalise the findings to all schools.

Claims can be tentatively made about external validity by comparing the attainment of schools that were passive participants. These are the schools that did not volunteer for the project but which were clustered with those that did. Passively participating schools in both the 'data' and 'data + conversation' treatment arms received performance data whereas passively participating control schools did not.

Table 7 presents cross-sectional treatment effects for passively participating schools. The effect for maths in 2015 achieves statistical significance, which confirms that the overall effect noted in Table 5 is driven by passively participating schools. The effect for science in 2015 is also borderline significant.

Table 7 Cross-sectional effects of treatment on attainment, passively participating schools

Year	Subject	Diff	SE	t-value	p-value
2014	Maths	0.27	0.20	1.40	0.16
2014	English	0.15	0.23	0.65	0.52
2014	Science	0.19	0.23	0.86	0.39
2015	Maths	0.42	0.19	2.26	0.02
2015	English	0.27	0.23	1.16	0.24
2015	Science	0.40	0.20	1.96	0.05

Source: Performance data on passively participating schools (*N* = 2,862)

Although clusters of schools were randomly assigned to treatment and control groups, the sample was balanced according to the characteristics of actively participating schools. Therefore, observed effects related to passively participating schools may be driven by pre-treatment differences between treatment and control schools. For instance, treatment schools may have already been higher attaining.

To circumvent this problem, difference-in-difference estimates for passively participating schools are presented in Table 8. The effects shown relate to treatment schools in both years following the intervention. All results are not significant.

Table 8 Difference-in-difference effects of treatment on attainment, passively participating schools

	Diff	SE	t-value	p-value
Maths	0.16	0.10	1.59	0.11
English	0.05	0.15	0.35	0.72
Science	0.20	0.16	1.26	0.21

Source: Passively participating schools (*N* = 2,862)

5.3. Changes in entry patterns

Using the difference-in-differences model, estimates are presented of the effects of the treatment on changes in:

- % of pupils entered for two or more GCSEs in sciences (EBacc definition)
- % of pupils entered for GCSEs in biology, chemistry and physics
- % pupils entered for GCSE English literature

The subject choices of the 2014 year 11 cohort would have already been made when the treatment was administered in November 2013. A treatment effect is therefore calculated based on data for 2015 only.

Table 9 shows a relatively small increase in entries in science among actively participating treatment schools, with the increase in entries in triple sciences achieving borderline statistical significance. Closer inspection reveals that this difference has arisen due to a fall in entries among control schools rather than an increase in entries among treatment schools.

There are no effects of treatment among passively participating schools (Table 10).

Table 9 Difference-in-difference effects of treatment on entries, actively participating schools

	Diff	SE	t-value	p-value
2 sciences	1.7%	2.6%	0.67	0.50
Biology, chemistry & physics	3.7%	1.9%	2.00	0.05
English literature	-1.9%	3.5%	-0.53	0.59

Source: Actively participating schools (*N* = 118)

Table 10 Difference-in-difference effects of treatment on entries, passively participating schools

	Diff	SE	t-value	p-value
2 sciences	-0.3%	1.2%	-0.26	0.80
Biology, chemistry & physics	0.5%	0.9%	0.55	0.58
English literature	0.3%	1.5%	0.19	0.85

Source: Passively participating schools (*N* = 2,862)

6. Conclusion

The findings from this research can assist policy makers in deciding which types of information should be provided to schools. In general, heads of department are confident about their ability to interpret and use pupil performance data. They are also very familiar with performance data, including information comparing their department with others in the same school, and with equivalent departments in other schools. However, departmental heads tend to overestimate the ability and performance of their department relative to others. This tendency to overestimate performance could be kept in check by encouraging heads of department to collectively review each other's exam performance. An annual inter-school meeting to analyse results would be a start. The Families of Schools database, which groups schools with similar intakes, can help make these comparisons more transparent. But what really matters is that performance is reviewed by respected, knowledgeable colleagues.

Providing schools with performance data and school exam entry information in alternative formats is useful to schools, particularly when it is clear and simple. However, providing information to schools is not sufficient to improve attainment outcomes or change entry patterns. Whilst schools are able to understand the information given to them, using it to aid decision-making and to implement change is more difficult.

This new evidence on the extent and nature of school collaboration in the UK suggests that it is difficult to foster. Schools struggle to collaborate because of time constraints and workload pressures, a lack of clarity about the purpose of collaboration projects and being unable to make contact with counterparts in other schools. Heads of departments aren't always confident that collaboration has a positive impact on their teachers and their department. However, they are keen to know about the exam board choices and entry profiles of similar schools, and high performing departments also tended to be encouraged to develop relationships with other schools by senior leadership teams. This indicates that there is an appetite for knowledge sharing and that senior leadership teams influence the degree of collaboration taking place. These findings should be of use in developing and refining policies which try to bring about school improvement through joint working arrangements, such as Teaching School networks and Specialist Leaders of Education.

7. References

Abrams, L. C., Cross, R., Lesser, E., & Levin, D. Z. (2003). Nurturing interpersonal trust in knowledge-sharing networks. *The Academy of Management Executive*, 17(4), 64-77.

Atkinson, M., Springate, I., Johnson, F., & Halsey, K. (2007). *Inter-School Collaboration: A Literature Review*. National Foundation for Educational Research.

Bell, M., Jopling, M., Cordingley, P., Firth, A., King, E. & Mitchell, H, (2006). *What is the impact on pupils of networks that include at least three schools? What additional benefits are there for practitioners, organisations and the communities they serve?* National College for School Leadership.

Brown, C. & Rogers, S (2014). Knowledge creation as an approach to facilitating evidence informed practice: Examining ways to measure the success of using this method with early years practitioners in Camden (London). *Journal of Educational Change*, 16(1), 79-99.

Brown, D. (2012). Understanding the better than average effect: motives (still) matter. *Personality and Social Psychology Bulletin*, 38(2), 209-219.

Burgess, S. (2014). *Understanding the success of London's schools*. CMPO Working Paper No. 14/333.

Chapman, C., & Hadfield, M. (2010). Realising the potential of school-based networks. *Educational Research*, 52(3), 309-32.

Coburn, C. E., & Talbert, J. E. (2006). Conceptions of evidence use in school districts: Mapping the terrain. *American Journal of Education, 112*(4), 469-495.

Coburn, C. E., & Turner, E. O. (2011): Research on data use: A framework and analysis. *Measurement: Interdisciplinary Research and Perspectives*, *9*(4), 173-206.

Dearden, L., Miranda, A., & Rabe-Hesketh, S. (2011). Measuring School Value Added with Administrative Data: The Problem of Missing Variables. *Fiscal Studies*, 32(2), 263-278.

DfE (2014) Teacher's Workload Survey 2013.

DfE (2015) Subject time series tables: SFR02/2015.

Dunning, D., Heath, C., & Suis, J. (2004). Flawed Self-Assessment: Implications for Health, Education, and the Workplace. *Psychological Science in the Public Interest*, 5(3), 69-106.

Farley-Ripple, E.N., & Buttram, J. (2015). The Development of Capacity for Data Use: The Role of Teacher Networks in an Elementary School. *Teachers College Record*, 117(4), 1-34.

Honig, M.I., & Coburn, C. (2008). Evidence-based decision-making in school district central offices: Toward a policy and research agenda*. Educational Policy, 22*(4), 578-608.

Ingram, D., Louis, K. S., & Schroeder, R. G. (2004). Accountability policies and teacher decision making: Barriers to the use of data to improve practice. *Teachers College Record*, 106(6), 1258–1287.

Inkpen, A. C., & Tsang, E. W. (2005). Social capital, networks, and knowledge transfer. *Academy of Management Review*, 30(1), 146-165

Jeffreys, H. (1961). The Theory of Probability (3rd ed.). Oxford. p. 432.

Jimerson, J.B. (2015). How are we Approaching Data-Informed Practice? Development of the *Survey of Data Use & Professional Learning*. *Educational Assessment, Evaluation and Accountability.* 1-27.

Kane, A. A., Argote, L., & Levine, J. M. (2005). Knowledge transfer between groups via personnel rotation: Effects of social identity and knowledge quality. *Organizational Behavior and Human Decision Processes*, 96(1), 56-71

Kane, A. A. (2010). Unlocking knowledge transfer potential: Knowledge demonstrability and superordinate social identity. *Organization Science*, 21(3), 643-660.

Kerr, K. A., Marsh, J. A., Ikemoto, G. S., Darilek, H., & Barney, H. (2006). Strategies to promote data use for instructional improvements: actions, outcomes, and lessons from three urban districts. *American Journal of Education*, 112, 496–520.

Leana, C. R., & Pil, F. K. (2006). Social capital and organizational performance: Evidence from urban public schools. *Organization Science*, 17(3), 353-366.

Light, D., Honey, M., Henze, J., Brunner, C., et al. (2005). *Linking data and learning: The Grow Network study*. New York: Education Development Center, Center for Children and Technology.

Means, B., Padilla, C., DeBarger, A., & Bakia, M. (2009). *Implementing data-informed decision making in schools: Teacher access, supports and use*. Report prepared for U.S. Department of Education, Office of Planning, Evaluation and Policy Development. Prepared by SRI International, Menlo Park, CA.

Meier, K. J., & O'Toole, L. J. (2001). Managerial strategies and behavior in networks: A model with evidence from US public education. *Journal of Public Administration Research and Theory*, 11(3), 271-294

Muijs, D., West, M., & Ainscow, M. (2010). Why network? Theoretical perspectives on networking. *School Effectiveness and School Improvement*, 21(1), 5-26.

Phelps, C., Heidl, R., & Wadhwa, A. (2012). Knowledge, networks, and knowledge networks a review and research agenda. *Journal of Management*, 38(4), 1115-1166

Schildkamp, K., & Kuiper, W. (2010). Data informed curriculum reform: Which data, what purposes, and promoting and hindering factors. *Teaching and Teacher Education*, *26*(3), 482-496.

Sims, S. (2012) *The development of quasi-markets in secondary education*. Institute for Government.

Stecker, P. M., Fuchs, L. S., & Fuchs, D. (2005). Using curriculum-based measurement to improve pupil achievement: review of research. *Psychology in the Schools, 42*(8), 795–819.

Sturgis, P., Smith, P., Hughes, G. (2006) A study of raising response rates in school surveys. DFES, RB721.

Sutherland, S. (2004). Creating a culture of data use for continuous improvement: a case study of an Edison project school. *The American Journal of Evaluation*, 25(3), 277–293.

Van Wijk, R., Jansen, J. J., & Lyles, M. A. (2008). Inter-and intra-organizational knowledge transfer: a meta-analytic review and assessment of its antecedents and consequences. *Journal of Management Studies*, 45(4), 830-853

Wohlstetter, P., Malloy, C. L., Chau, D., & Polhemus, J. L. (2003). Improving schools through networks: A new approach to urban school reform. *Educational Policy*, 17(4), 399-430.

Wohlstetter, P., Smith, J., & Gallagher, A. (2013). New York City's Children First Networks: turning accountability on its head. *Journal of Educational Administration*, 51(4), 528-549.

Appendix 1: creating the clusters of similar schools

Schools in Scope

All state-funded mainstream schools in England that met the following criteria were included in the sampling frame:

1) key stage 4 attainment data for the 2011/12 academic year
2) a January 2013 school census return, with at least 60 pupils in at least one of year 9, year 10 or year 11
3) not listed in Edubase to close by 31/08/2014

Where appropriate, recently opened schools have been linked to predecessor schools. In total 2,980 schools are in scope.

Variables used in clustering

The school sex of intake is used to partition clusters of schools, thus creating clusters of boys' schools, clusters of girls' schools and clusters of mixed schools.

Within each partition, clusters of geographically proximal and similar schools were created on the basis of:

a) geographic distance (as-the-crow-flies)
b) estimated mean GCSE grade in mathematics (for years 9,10 and 11 separately), derived from the pupil's key stage 2 prior attainment, pupil contextual factors (e.g. SEN, FSM, gender) and school contextual factors (e.g. % FSM, mean prior attainment)
c) variance in GCSE grade in mathematics (for years 9,10 and 11 separately)
d) % pupils with a first language other than English (for years 9,10 and 11 combined)

All factors were standardised (using school-level means and standard deviations) prior to running the process. Each of these factors were weighted using the values set out in Table 11.

Table 11 Weighting of variables used to create clusters

	Single-sex	Mixed
Geography	5	6
Mean estimates	6	6
Variance in estimates	0.5	0.5
% first language not English	1	2
Total	12.5	14.5

Source: Schools in scope as per rules above (*N* = 2,980)

Acceptance criteria

There are two partially competing requirements:

- To include as many schools as possible, preferably all schools
- Ensuring each pair of schools within each cluster is sufficiently ‘close’, both geographically and statistically

Two acceptance criteria are therefore imposed for each pair of schools within a cluster:

- Distance <=80km
- two or more of the three maths estimates for years 9, 10 and 11 are within half a GCSE grade

The second criterion was relaxed slightly for schools in the top (or bottom) 10% of schools nationally. The estimates are also ‘shuffled’ such that two schools with identical estimates for year 10 but reversed estimates for years 9 and 11 are considered acceptably close.

The clustering process begins by identifying all pairs of schools nationally that meet the acceptance criteria above. There are 75 schools that are not sufficiently ‘close’ to at least five other schools and hence they are not included in the clustering process.

Clustering

A standard k-means approach is adopted. Broadly, the standard algorithm works as follows:

1. Initialise *k* points $\underline{C}$ within the working space, typically by
   1.1. allocating all observations to *k* clusters and taking the averages for these clusters as the initial centre points; or
   1.2. randomly selecting *k* observations and using these points.
2. Assign each observation to the member of $\underline{C}$ it is “closest” to.
3. Recalculate the centre points $\underline{C}$ as the averages for the observations assigned to them.
4. If the set $\underline{C}$ is unchanged, break; repeat from step 2 otherwise.

Advantages of such an approach to this application include:

1) it is computationally simple, meaning that convergence is typically fast, making it possible to very quickly try different options for definitions of proximity (e.g. experimenting with different weightings);
2) it is relatively easy to understand;
3) the clusters produced are mutually exclusive and exhaustive, as required.

Initialisation

An initial set of clusters (containing six schools per cluster) was produced by ranking all schools-in-scope on their average maths estimate across years 9, 10 and 11.

Proximity

This application uses a sum of squared differences between values (of the weighted variables listed above) of the school and each of the cluster centre points to determine which cluster each school should be assigned to.

Assignment

A typical k-means algorithm assigns observations to clusters by simply selecting the cluster that minimises the squared (standardised) distance or, to put it more simply, by selecting the closest centre point. The requirement, however, is to ensure that each cluster is comprised of exactly six members.

To that end, assignment uses the following algorithm:

- For each observation, *o*, not already assigned to a cluster, find the closest centre point, *p*, for clusters that have fewer than six members.
- Assign *o* to the cluster associated with *p* where it is at most the (6-*m*)th closest to *p*, where *m* is the number of observations already assigned to the cluster.
- Repeat from 1 until no clusters have fewer than six members.

The remaining observations are those most distant from the clusters and can be considered outliers; they occur when the population is not divisible by six.

Adjustment

The process described above by definition produces an optimal national solution: the set of possible clusters of six schools that minimises the sum of squared standardised differences between each school and its cluster centroid.

Next, clusters were identified in which a single school was not sufficiently close to the other five. Attempts were made to find a replacement school (such that the acceptance criteria are satisfied) from the set of unallocated schools.

2,634 (88%) of schools were allocated to clusters by the end of this stage.

Stage 2 Clusters

Attempts were then made to create clusters from the remaining unallocated schools by identifying sets of six schools that reciprocally meet the acceptance criteria.

Another 66 schools were allocated to clusters by the end of this stage.

Stage 3 Clusters

Finally, the acceptance criteria was removed and clusters were created from the remaining 282 schools. After manual inspection, another 66 schools were deemed fit for use in the project.

In total, 36 clusters (216 schools) were deemed unfit for use due to geographical and/ or statistical distances between pairs of schools.

Appendix 2: Survey response rates

Baseline survey

The response rate for the baseline survey was just under 64%, with comparable response rates for English, maths and science departments. One respondent failed to state which department they worked in.

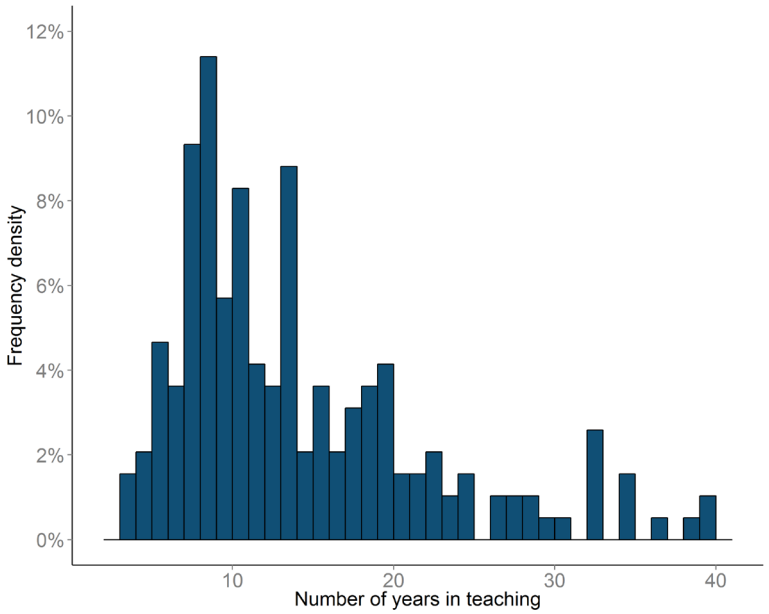
Table 12 Response rates for the Baseline survey

	Overall	English	Maths	Science
Frequency	226	74	79	72
Percent	63.8%	62.7%	66.9%	61.0%

Source: Baseline survey (*N* = 226)

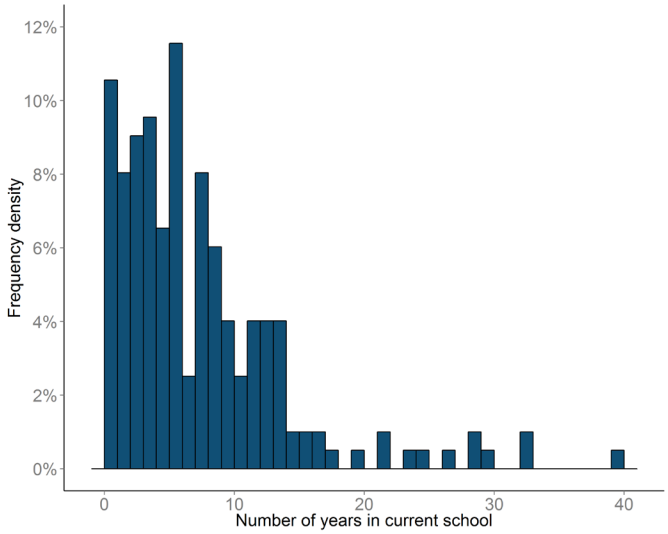
The respondents to the survey were overwhelmingly Heads of Department (98%) with just four Deputy Heads and two key stage leaders. On average, they had been teaching for 13.8 years (none had been teaching for less than three years), had been working at their current school for 7 years, and had been in their current job for 3.9 years. Figure 28, Figure 29 and Figure 30 show the full distribution of responses.

Figure 28: Respondents' years in profession



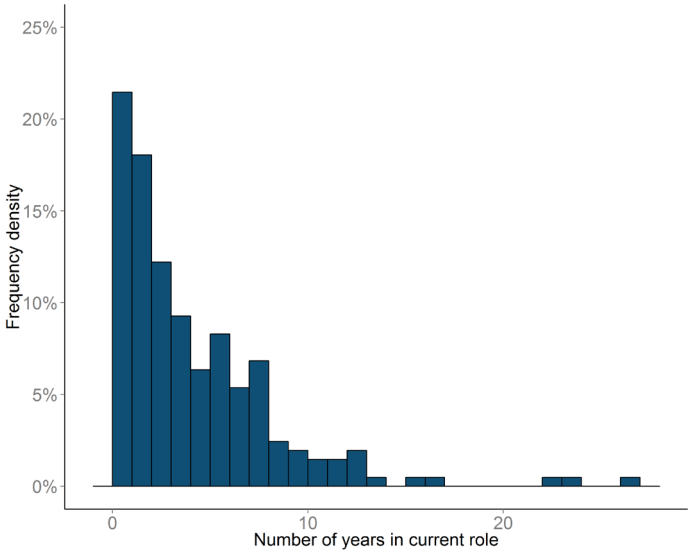
Source: Baseline survey (*N* = 193)

Figure 29: Respondents' experience at current school



Source: Baseline survey (*N* = 199)

Figure 30: Respondents' years in current role



Source: Baseline survey (*N* = 205)

Data survey

The response rate for the data survey was just over 42% (91 partial/full responses from 216 invitations) with better response rates from maths departments than for English and science departments.

Table 13 Response rates for the Data survey

	Overall	English	Maths	Science
Frequency	91	29	34	28

	Overall	English	Maths	Science
Percent	42.1%	40.3%	47.2%	38.9%

Source: Data survey (*N* = 91)

Appendix 3: Surveys and data information sent to schools

Baseline survey of Subject Excellence Clusters

Your role in the school

1) My department teaches:

2) My school name is:

3) Please select the job title that describes your role in your department best:

4) Enter the number of month and year since you...
…began in this role
…started teaching in this school
…started your first teaching job

Using data to set targets

5) Does your department set GCSE target grades for pupils (whether communicated to pupils or not)?

6) How do you set targets for your pupils?

7) Do you adjust any child's target to reflect personal views of the child's capabilities in the subject?

8) Who is most responsible for setting pupil targets in your department?

Using data to set pupil targets

9) How often does your department collect teacher assessments of pupil progress or predictions of pupils' likely grades?

10) What are any teacher predicted grades or assessments of progress that you routinely collect used for?

Using data to evaluate performance

11) Does your department use data to evaluate your annual GCSE results?

12) How do you evaluate your GCSE results?

Your department's performance in recent years

13) How well did your year 11 pupils perform in GCSE exams in your subject in 2012 and 2013?

14) Looking back to last year's year 11 performance in GCSE exams, is there a set of pupils who performed particularly well in your subject?

Teacher appraisal and performance

15) As a Head of Department or Deputy Head of Department, could you confidently name your most effective teacher(s)? (We will not ask for their name!)

16) If yes, how do you know they are effective?

Setting goals for teachers

17) Do you set explicit overall class targets for individual teachers (e.g. the teacher should aim to achieve 60% A*-C within the class)?

18) If so, who sets these targets and how do they do it?

19) How important is the use of pupil achievement data for teacher appraisal in your department?

20) What type of explicit goals are teachers in your department given each year?

Your contact with other schools

21) Does the Senior Leadership Team at your school encourage you to cultivate relationships with other schools or departments?

22) Have you had contact with teachers at other schools (in a professional capacity) in the past 12 months?

23) If yes, please describe the nature of this contact

24) Has anyone in your department consciously decided to work with teachers at another school to improve any aspect of how your department works?

Your contact with other schools

25) If yes, how did you choose which schools to work with?

26) Describe what types of interactions and discussions you had with the teachers at other schools

27) Can you name a school that your department makes regular contact with, either in person or via telephone? If no, then leave blank.

Data survey of Subject Excellence Clusters

Your department's performance data

1) Find your school's performance information on the first page of the PDF. Which groups of pupils performed statistically better or worse compared to schools like yours in the rest of the country?

2) Look at the performance of your school's group of pupils with level 4A prior attainment at KS2. Which one of these statements best describes their performance at GCSE?

3) In general is this data on page 1 clear to you? Is there anything you don't really understand?

4) Did you learn anything new about your department's 2013 GCSE performance from the information on page 1?

Your department's pupil entry profile data

5) Find your school's pupil entry information on the second page of the PDF. Did you find this data useful? Does it help you reflect on your school's entry profile choices in 2013?

6) Are there other details about entry choices of similar local schools that you would find useful?

Your school's performance data

7) Find your school's general performance information on the last page of the PDF. Looking at the information on the GCSE performance of your school's core departments over the past three years, does this information reflect your understanding of these department's effectiveness? Why or why not?

8) In general, what other data on department or school performance would you find helpful?

Your knowledge of local schools

9) Below is the list of your 10 nearest secondary schools (that had KS4 performance data in 2013). Tick any statements that apply to each school in turn. "This school...":
…has a similar intake to my school
…is one in which I have close professional relationships with some of the staff
…is high performing
…is part of the same federation or chain or teaching school alliance as my school
…has a similar approach to teaching and learning as my school

10) During the last three years, has anyone in your department collaborated with teachers in the equivalent department in the following schools?

Interactions with local schools

11) Which types of collaborative activities have you engaged in with other schools over the last three years? Please tick all that apply.

Interactions with local schools

Please indicate whether you agree or disagree with the following statements about the effects of collaborations you have experienced on teachers in your department:

12) Collaborating with other schools has increased our teachers' engagement with their work

13) Collaborating with other schools has increased our teachers' morale at work

14) Collaborating with other schools has increased our teachers' confidence at work

15) Collaborating with other schools has improved our teachers' pedagogical skills

16) Collaborating with other schools has improved our department's lesson resources

---

Your ESEC cluster schools

17) Turning to the five other schools that we gave you performance information on, has anyone in your department had contact with teachers in the equivalent department in the following schools?

18) Do you plan to contact departments in any of these schools during the remainder of the academic year?

19) Which of these do you think are credible comparators in terms of similarity of pupil home backgrounds and alignment of departmental needs/interests?

20) Returning to the first page of the PDF we sent you. For each of the schools we gave you data on, if they performed significantly better or worse than your department in the 2013 GCSE performance tables, do you know why?

Your department's plans

21) Are you doing any of the following?
- We are reviewing our choice of exam boards or specifications at KS4
- We are reviewing our policies regarding early entry for GCSE
- We are reviewing our department's Year 9 curriculum
- We are reviewing the number of hours of teaching time in our subject
- We are reviewing the development goals of our department
- We are reviewing how we allocate teachers to classes
- We are making substantial changes in our schemes of work

22) Has the performance information data that we sent you influenced any of the other decisions? If so, state how.

Example performance data report

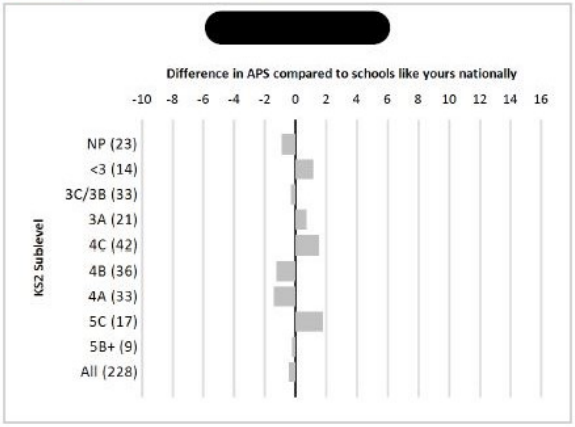
Key Stage 4 Average Point Score 2013 -

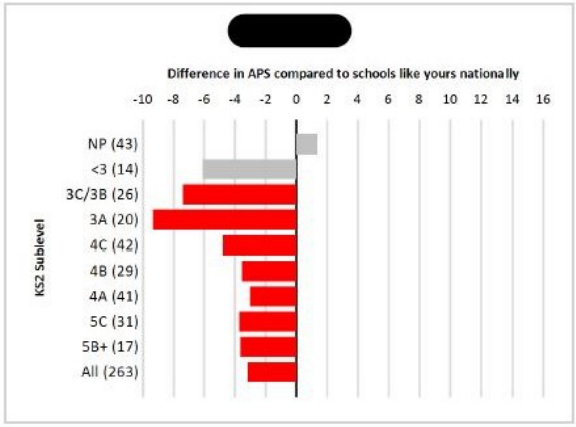
The charts below summarise attainment in Mathematics in 2013 at the six schools in your cluster. They can be used to identify particular groups of pupils based on prior attainment that achieved more or less well at schools in your cluster than similar pupils in similar schools nationally.

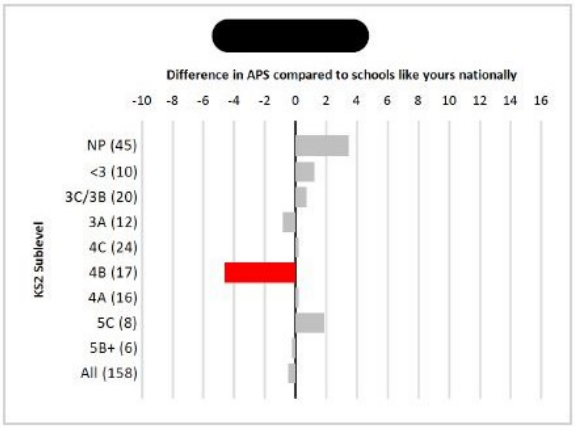
On each chart, the solid black line denotes the average point score (APS) of schools like yours. The bars represent differences in APS compared to schools like yours nationally for groups of pupils based on their end of Key Stage 2 sublevel in Mathematics. The colours of the bars indicate whether the difference to the average of schools like yours nationally is statistically significant. The numbers in brackets on the vertical axis denotes the number of pupils in each prior attainment band. Attainment data for 2013 is validated, i.e. has been checked by schools prior to publication of Performance Tables.

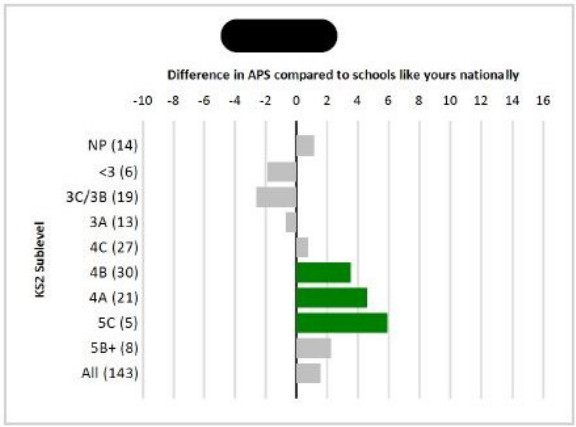
Key

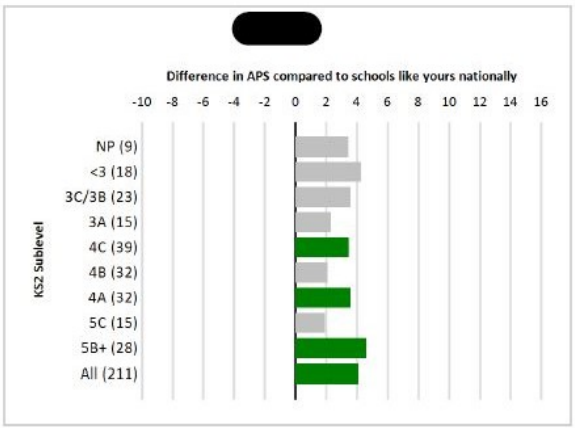
- Significantly higher than the average of schools like yours nationally
- Not significantly different from the average of schools like yours nationally
- Significantly lower than the average of schools like yours nationally

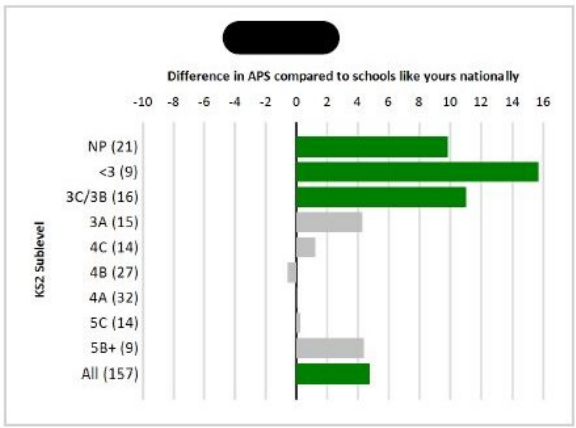












Entries in 2013

This page shows the different types of Mathematics qualification entered by pupils who reached the end of Key Stage 4 in 2013. Prior attainment is based on students' end of Key Stage 2 sublevels in Mathematics. Significance tests are used to highlight entry rates that are significantly different (higher or lower) from the national average. These tests should not imply that any particular type of entry is more or less appropriate given the prior attainment of students. The data is presented to inform your professional judgment about the most suitable types of qualifications in which to enter your students.

Key

Rate is significantly higher than schools like yours nationally

Rate is significantly lower than schools like yours nationally

School		% entered for an AS level in maths	% entered early for GCSE maths	% entered twice or more in GCSE maths	% entered for GCSE maths	% entered for level 2 qualification in numeracy	% entered for level 1 qualification in numerary	Pupils
[redacted]	No prior attainment	0%	91%	87%	96%	30%	0%	23
	Below Level 3	0%	93%	86%	93%	7%	0%	14
	Level 3B / 3C	0%	94%	91%	94%	21%	0%	33
	Level 3A	0%	100%	100%	100%	43%	0%	21
	Level 4C	0%	95%	81%	100%	17%	0%	42
	Level 4B	0%	94%	78%	94%	36%	0%	36
	Level 4A	0%	97%	85%	97%	33%	0%	33
	Level 5C	0%	100%	82%	100%	47%	0%	17
	Level 5B+	0%	100%	89%	100%	44%	0%	9
	All	0%	96%	86%	97%	29%	0%	228
[redacted]	No prior attainment	0%	98%	95%	100%	0%	0%	43
	Below Level 3	0%	86%	86%	100%	0%	7%	14
	Level 3B / 3C	0%	81%	77%	92%	0%	8%	26
	Level 3A	0%	95%	95%	95%	0%	0%	20
	Level 4C	0%	86%	81%	90%	2%	2%	42
	Level 4B	0%	100%	97%	100%	0%	0%	29
	Level 4A	0%	95%	95%	95%	2%	2%	41
	Level 5C	0%	100%	94%	100%	3%	0%	31
	Level 5B+	0%	100%	94%	100%	0%	0%	17
	All	0%	94%	90%	97%	1%	2%	263
[redacted]	No prior attainment	0%	100%	93%	100%	27%	0%	45
	Below Level 3	0%	90%	90%	90%	20%	0%	10
	Level 3B / 3C	0%	95%	90%	95%	35%	10%	20
	Level 3A	0%	100%	92%	100%	50%	0%	12
	Level 4C	0%	100%	88%	100%	50%	13%	24
	Level 4B	0%	82%	76%	88%	65%	12%	17
	Level 4A	0%	100%	100%	100%	69%	6%	16
	Level 5C	0%	100%	100%	100%	88%	0%	8
	Level 5B+	0%	100%	83%	100%	83%	0%	6
	All	0%	97%	91%	97%	46%	5%	158
[redacted]	No prior attainment	0%	0%	0%	100%	0%	0%	14
	Below Level 3	0%	0%	0%	100%	0%	0%	6
	Level 3B / 3C	0%	0%	0%	100%	0%	0%	19
	Level 3A	0%	8%	8%	100%	0%	0%	13
	Level 4C	0%	0%	0%	100%	0%	0%	27
	Level 4B	0%	0%	0%	100%	0%	0%	30
	Level 4A	0%	0%	0%	100%	0%	0%	21
	Level 5C	0%	0%	0%	100%	0%	0%	5
	Level 5B+	0%	0%	0%	100%	0%	0%	8
	All	0%	1%	1%	100%	0%	0%	143
[redacted]	No prior attainment	0%	89%	67%	89%	0%	0%	9
	Below Level 3	0%	100%	100%	100%	0%	0%	18
	Level 3B / 3C	0%	100%	100%	100%	0%	0%	23
	Level 3A	0%	100%	100%	100%	0%	0%	15
	Level 4C	0%	100%	92%	100%	0%	0%	39
	Level 4B	0%	100%	94%	100%	0%	0%	32
	Level 4A	0%	100%	75%	100%	0%	0%	32
	Level 5C	0%	100%	73%	100%	0%	0%	15
	Level 5B+	0%	100%	57%	100%	0%	0%	28
	All	0%	100%	85%	100%	0%	0%	211
[redacted]	No prior attainment	0%	100%	100%	100%	33%	67%	21
	Below Level 3	0%	89%	89%	89%	22%	100%	9
	Level 3B / 3C	0%	100%	100%	100%	13%	94%	16
	Level 3A	0%	100%	87%	100%	7%	87%	15
	Level 4C	0%	93%	86%	93%	29%	86%	14
	Level 4B	0%	96%	89%	96%	48%	96%	27
	Level 4A	0%	100%	94%	100%	63%	94%	32
	Level 5C	0%	100%	100%	100%	64%	86%	14
	Level 5B+	0%	100%	100%	100%	56%	78%	9
	All	0%	98%	94%	98%	40%	88%	157

School Context Indicators

Based on January 2013 School Census. 'Schools like yours' are 30 other schools nationally with similar pupil intakes.

Pupil Characteristics

Indicator	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Number on roll years 7 to 11	877	908	590	782	1059	888	25367	
% eligible for free school meals	41%	36%	26%	56%	41%	41%	31%	16%
% ever eligible for free school meals (last 6 years)	66%	50%	47%	74%	64%	63%	49%	28%
% first langauge not English	39%	35%	39%	34%	38%	39%	36%	14%
% boys	60%	49%	52%	72%	58%	66%	53%	51%
% SEN (School Action Plus or Statement)	16%	7%	15%	16%	25%	19%	14%	9%

Ethnicity

Indicator	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
% White British	18%	49%	41%	38%	31%	22%	50%	76%
% White Irish	<1%	<1%	<1%	1%	<1%	<1%	<1%	<1%
% White Other	10%	12%	18%	14%	15%	9%	11%	4%
% Mixed White/Black Caribbean	6%	1%	4%	5%	2%	5%	3%	1%
% Mixed White/Black African	2%	3%	<1%	2%	4%	1%	<1%	<1%
% Mixed White/Asian	<1%	<1%	<1%	<1%	2%	<1%	1%	<1%
% Mixed (other)	3%	1%	1%	5%	3%	3%	2%	1%
% Indian	<1%	1%	1%	<1%	3%	0%	4%	2%
% Pakistani	<1%	2%	4%	<1%	3%	<1%	12%	4%
% Bangladeshi	1%	5%	<1%	4%	<1%	4%	2%	1%
% Asian (other)	3%	2%	7%	<1%	4%	1%	2%	1%
% Black Caribbean	14%	2%	6%	7%	2%	9%	3%	1%
% Black African	24%	17%	11%	14%	16%	30%	7%	3%
% Black (other)	4%	3%	2%	2%	2%	4%	<1%	<1%
% Chinese	5%	<1%	0%	<1%	<1%	3%	<1%	<1%
% Any other group	8%	2%	2%	6%	11%	8%	2%	1%
% Unknown ethnicity	3%	4%	<1%	<1%	<1%	<1%	<1%	1%

Prior Attainment

Indicator	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
KS2 average point score (year 7)	27.1	25.3	25.8	26.3	26.0	27.4	26.3	28.1
KS2 average point score (year 8)	26.7	25.0	25.1	26.3	26.3	26.7	25.8	27.6
KS2 average point score (year 9)	25.4	25.3	25.6	26.2	25.6	26.3	25.7	27.6
KS2 average point score (year 10)	25.8	26.7	26.0	26.2	26.4	26.5	25.9	27.8
KS2 average point score (year 11)	26.0	26.7	25.5	26.3	26.2	26.1	26.0	27.8
% pupils with high prior attainment (year 7)	13%	10%	3%	6%	5%	11%	8%	20%
% pupils with high prior attainment (year 8)	13%	5%	5%	8%	7%	9%	8%	20%
% pupils with high prior attainment (year 9)	7%	9%	3%	7%	7%	10%	8%	20%
% pupils with high prior attainment (year 10)	7%	7%	5%	3%	11%	8%	8%	20%
% pupils with high prior attainment (year 11)	6%	8%	4%	6%	11%	5%	8%	20%
% pupils with low prior attainment (year 7)	22%	42%	41%	32%	33%	23%	33%	19%
% pupils with low prior attainment (year 8)	23%	47%	39%	27%	30%	23%	33%	19%
% pupils with low prior attainment (year 9)	36%	36%	32%	28%	37%	32%	33%	19%
% pupils with low prior attainment (year 10)	38%	27%	32%	31%	27%	27%	32%	19%
% pupils with low prior attainment (year 11)	30%	29%	39%	28%	33%	28%	32%	19%
% pupils without prior attainment	13%	18%	18%	5%	5%	10%	9%	3%

Key Stage 4 Headline Indicators

This report presents headline indicators for schools in your cluster for the last three academic years. Shading denotes performance that is significantly different from the average of 'Schools like yours'. These are 30 other schools nationally with similar pupil intakes.

Key

Rate is significantly higher than schools like yours nationally

Rate is significantly lower than schools like yours nationally

Overall

Indicator	Year	[redacted]							
Number of pupils	2011	219	230	206	109	208	167		
	2012	223	185	146	144	206	156		
	2013	228	263	158	143	211	157		
% English Baccalaureate	2011	8%	0%	0%	5%	4%	4%	2%	16%
	2012	2%	2%	0%	1%	5%	10%	3%	16%
	2013	15%	6%	6%	11%	20%	10%	9%	23%
% 5 A*-A	2011	16%	11%	15%	17%	18%	24%	13%	22%
	2012	8%	6%	10%	12%	19%	10%	15%	23%
	2013	10%	4%	8%	11%	22%	7%	15%	21%
% 5 A*-C incl. English & maths	2011	48%	47%	42%	36%	53%	69%	43%	59%
	2012	47%	43%	47%	60%	54%	60%	47%	60%
	2013	52%	44%	56%	46%	67%	57%	50%	62%
Capped point score (with equivalents)	2011	291.2	304.1	323.4	327.8	336.5	326.5	322.4	344.0
	2012	289.6	287.9	313.5	334.8	328.7	298.3	329.4	348.7
	2013	317.0	296.0	303.6	342.7	340.3	306.6	332.6	347.6
Average grade per student (GCSE only)	2011	34.9	32.4	31.3	34.7	35.0	38.5	33.6	38.6
	2012	35.3	33.5	31.3	37.4	35.5	37.8	33.8	38.8
	2013	35.3	32.3	33.4	35.5	37.7	36.7	34.3	38.7

Core Subjects

Indicator	Year	[redacted]							
Average point score in English	2011	38.2	35.9	32.6	35.7	37.3	38.3	35.4	39.8
	2012	37.0	35.2	33.4	37.3	37.8	37.5	36.0	39.6
	2013	38.1	34.6	34.6	36.0	40.8	39.9	37.1	39.9
Average point score in mathematics	2011	36.2	33.3	34.6	36.3	37.5	40.9	34.4	39.0
	2012	33.7	33.7	36.6	40.1	37.5	41.4	35.2	39.7
	2013	35.5	32.8	35.5	37.5	40.0	40.7	35.9	39.9
Average point score in science (pupils entered for 2 GCSEs in science)	2011	40.7	37.0	42.4	43.5	40.3	42.4	40.0	43.4
	2012	40.1	40.0	40.4	42.6	35.7	42.8	40.1	43.4
	2013	35.8	39.3	46.7	43.1	43.7	38.8	39.4	42.7
% entered in mathematics	2011	95%	91%	97%	99%	100%	96%	98%	99%
	2012	95%	96%	99%	100%	100%	99%	98%	99%
	2013	97%	97%	97%	100%	100%	98%	99%	99%

[redacted]

© Education Datalab, FFT Education Ltd

Reference: DFE-RR511

ISBN: 978-1-78105-623-3

This research was commissioned under the under the 2010 to 2015 Conservative and Liberal Democrat coalition government. As a result the content may not reflect current Government policy. The views expressed in this report are the authors' and do not necessarily reflect those of the Department for Education.

Any enquiries regarding this publication should be sent to us at:
Laura.Pearce@education.gov.uk or www.education.gov.uk/contactus

This document is available for download at www.gov.uk/government/publications