Understanding school group efficiency

Working paper

Bobbie Mills

October 2021

EDUCATION POLICY INSTITUTE

Research Area:
School Performance, Admissions, and Capacity

This work was produced using statistical data from ONS. The use of the ONS statistical data in this work does not imply the endorsement of the ONS in relation to the interpretation or analysis of the statistical data. This work uses research datasets which may not exactly reproduce National Statistics aggregates.

Acknowledgements

Thanks to Professor Geraint Johnes and Professor Subhash Ray for their assistance in developing our understanding of the modelling design. In addition, our thanks to the members of the Effective School Group Advisory Board for their insightful contributions.

The analysis and interpretation presented in this paper is entirely the responsibility of EPI and not of our external associates.

Contents

1. Executive Summary

Significant public funding is invested in education in England each year. School leaders and other education bodies are responsible for deploying this resource efficiently and sustainably.

It is a technical challenge for researchers and policy makers to identify whether schools and trusts are run efficiently, and which of them have the most efficient operating models. By efficient, we mean achieving a certain level of output (e.g. pupil outcomes), with the lowest possible level of input (e.g. size of leadership team or experience of teachers). The difficulty in identifying efficiency arises firstly because spending patterns and operating models vary substantially across the different contexts of the school system. Secondly, there is no agreed measure for a good outcome for either schools or pupils.

Understanding efficiency is increasingly important in the context of real-terms cuts to pupil funding over the last decade, where school leaders have been under pressure to make increasing savings. The Covid-19 pandemic has brought further pressures to school finances.[1] Studying efficiency can help us to understand what genuine efficiency savings a school can make without compromising the quality and sustainability of education and its workforce, and where resources are potentially not being put to best use.

Efficiency is also ripe for examination because the proliferation of school groups in the last decade, mainly in the form of multi-academy trusts (MATs), is partly built on the theory that efficiency can be achieved through economies of scale.

This working paper presents the first stages in the development of our measures of efficiency at school group level. It outlines an approach that is based on Data Envelopment Analysis (DEA) and presents initial results for feedback and improvement via the consultation.

This is the first publication in a broader programme of work investigating the features of effective school groups. In tandem, we are also developing measures of sustainable workforce deployment and pupil inclusion to identify the school groups that operate most effectively to achieve positive outcomes for young people.

In addition, we will conduct in-depth qualitative research with teachers, leaders and local government and initiate a panel survey to gain insight into how schools work in groups across these domains to achieve the most positive outcomes.

This working paper covers:

- An overview of existing methods for studying school efficiency
- The development and selection of key "input" and "output" variables, which are designed to capture the operating model or "production function" of a school
- A discussion of how we can apply methods of efficiency estimation to different types of school groups, in order to compare across diverse groupings
- A step-by-step demonstration of data envelopment analysis (DEA) as a method for studying efficiency

[1] Bobbie Mills, "Assessing Covid-19 Cost Pressures on England's Schools" (Education Policy Institute, December 2020).

- A presentation and discussion of initial findings from our current DEA model

The DEA model presented in this paper identifies efficient schools with a range of performance profiles. That is to say that the model successfully identifies efficiency in schools that have both more and less challenging cohorts. This is because the model is designed to reward schools and trusts that achieve higher attainment than would be expected given the characteristics of their pupil intake, rather than simply rewarding schools with high attainment overall.

This is a foundational model which we want to develop further before drawing strong conclusions from our results. However, using the initial results from this model, we make tentative comparisons of the features of the most and least efficient schools and trusts. We find that:

- The most significant differences are the size of leadership team relative to pupil numbers, and lower spending on 'back office' functions as a proportion of total expenditure.
- Expenditure on education support staff is also comparatively lower among the most efficient schools and trusts, however some of the most efficient schools still have similar spends in this area compared with those with lower efficiency.
- In comparison with other input variables, there is striking overlap between the most and least efficient schools and trusts in terms of the level of teaching experience employed. Overall, it appears that high efficiency does not tend to be achieved with significantly less experienced teaching staff.

Information on how to contribute feedback is included at the end of this paper. Our planned next steps in developing these measures are to:

- Explore models that include multiple outputs, including 'negative' outputs such as unsustainably high staff turnover
- Further develop our understanding of alternative methods such as Stochastic Frontier Analysis
- Improve our inputs to better differentiate between school operating models
- Further develop our thinking on comparing across different types of school group
- Continue to ensure our model is as fair as possible in how it accounts for school context and challenge.

2. Introduction

In England, the government spends approximately £45 bn per year in revenue funding to schools.[2] This funding is deployed by leaders of schools, trusts, and other education authorities to deliver education to all young people nationally. The funding is used in conjunction with other additional resources, which also may be shared among groups and networks.

How limited resources can be put to best use depends on the context and outlook of decision-makers in education. There is no consensus on how best to deliver quality education within a sustainable organisation. For example, school leaders may decide to invest more in teaching assistants, depending on the needs of their pupils. Equally, fewer experienced qualified teachers may be employed, potentially due to a shortage in certain subjects, or potentially due to an express policy of hiring younger staff to develop within the organisation. Other decisions must be made around staffing and leadership development, including managing out those who underperform. Equally, on curriculum materials, some school groups may choose to invest in buying in or developing centralised materials with the view to delivering a consistent offer and freeing up teaching hours from preparation, whereas others may prefer to encourage teacher autonomy regarding curriculum materials. Opportunities and decisions also exist, in theory, for achieving economies of scale.

All in all, these are decisions to be made, executed, and monitored by school governance teams as to how resources will be used to achieve their aims. There is no single route to quality, sustainable, resource-efficient education and indeed there are multiple definitions of positive educational outcomes. England is a particularly diverse education system, comprising different delivery models with different “missions” for the kind of education to be delivered to young people, and there are differing constraints on this depending on school type.

The underlying theory of today’s education policy landscape is that improvement in education can be best achieved with effective school leaders working through groups and networks, particularly multi-academy trusts, and the nature of school groups in England adds an additional layer of diversity to the school system. The vast majority of pupils in England now attend a school that is part of a school group. Most schools do not operate in isolation but are part of a local authority, academy trust, federation, foundation trust, diocese or diocesan MAT. The importance of understanding the features of effective school groups has increased with the rapid expansion of the academies programme since 2010. Over three-quarters of secondary-aged pupils, and a third of primary-aged and special school pupils, are now being educated in academies and free schools. Over half of academies are in multi-academy trusts of six or more schools, and a significant number are in trusts that are even larger. The difference made by introducing academy trusts as key players in the education landscape in England is the greater potential to move between groups, and a greater diversity in how groups choose to operate. This dynamism, autonomy, and diversity is intended to have positive implications for pupil outcomes, particularly for the most disadvantaged, and to be one of the key drivers for school improvement in the system. Raising educational standards for the most disadvantaged was a keystone in the coalition government’s case for expanding the academies programme, but academies are not the only group with potential to benefit pupils through operating

[2] https://explore-education-statistics.service.gov.uk/find-statistics/school-funding-statistics

in numbers. A significant proportion of schools, particularly primary schools, remain outside the academy system. In fact, in 60 of the 151 local authority areas in England, less than a quarter of primary pupils are being educated in academies. Local authority schools, though not grouped by any express intention on the part of school leadership, may nevertheless be considered a loose group of schools within a local area. Our mixed system also encompasses schools in federations, foundation trusts, and those of religious denomination which have important relationships with their diocese. Equally there are standalone academies and free schools that are separate from any of these group-types.

A challenge then, for researchers and policymakers, is to understand which schools and school groups are most efficient with their resources, given that they can be deployed in so many ways and across so many types of school grouping. By "efficient", we mean achieving the highest possible "output" given a certain level of "input".[3] Successfully identifying the most efficient schools and school groups is an opportunity to learn how they get the most out of often-limited resources. Studying efficiency in the school system in England can also provide insight into where additional resources are most needed (and therefore should be best targeted), and where additional resources may not be being put to best use. Through this lens we can also explore where efficiency savings may in fact be cuts to quality.

This working paper is a first step in developing our measures of efficiency among schools and school groups in England. It outlines various methods for estimating efficiency and presents results for feedback and improvement.

Talking about efficiency can be a controversial subject in the context of recent years of real-terms cuts to pupil funding; increased pressure on teachers and leaders to deliver more as children's services are reduced in local government and elsewhere; pressure on school finances due to the Covid-19 pandemic[4]; pledges of increased education spending potentially being misplaced as increases to per pupil spending will not target the poorest schools[5]; and finally the recent history of teacher shortages, particularly in certain subjects, compounded by issues in teacher retention. School leaders point to trends of education centres being asked to do more with decreasing resources, whilst policymakers need a better understanding of where additional support can make most impact in the context of limited resources.

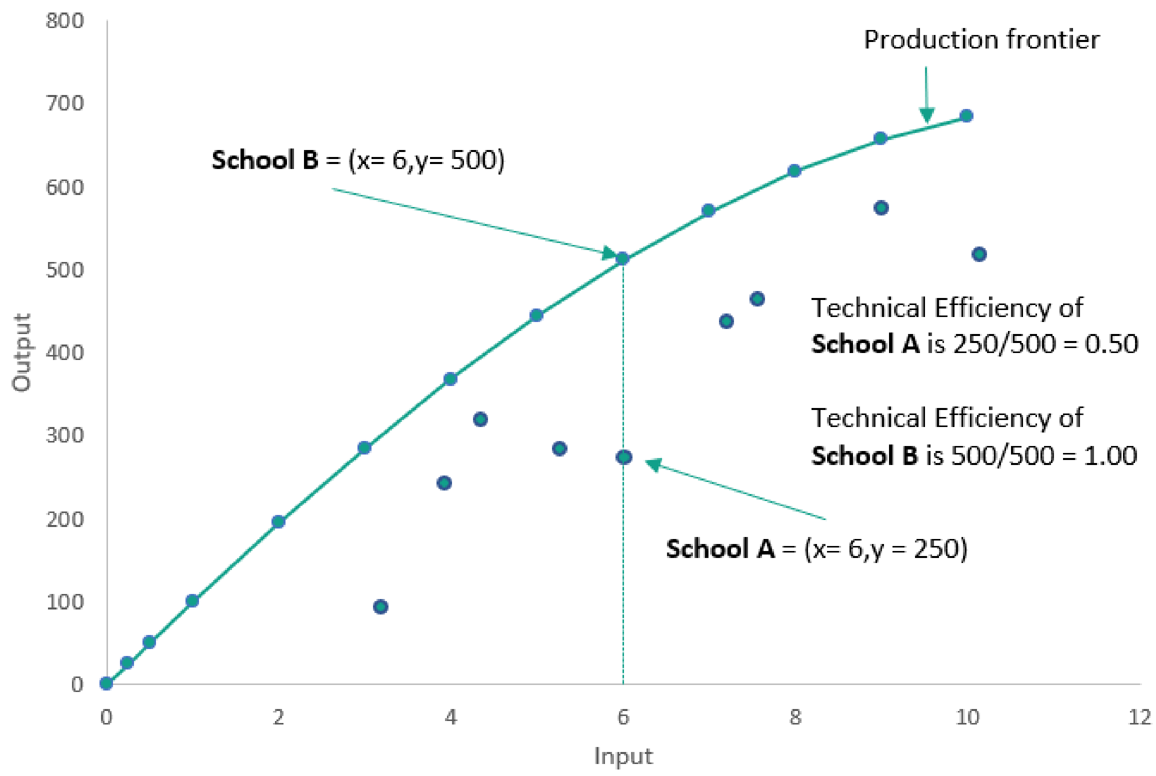
We approach efficiency as a lens for understanding how existing resources might be best used in the system, and as a route for identifying practices that bring genuine efficiencies without compromising workforce sustainability or quality of education (for instance in terms of breadth of curriculum). While the measures of efficiency reported in this paper do not achieve an entirely balanced and holistic view of efficiency, they should be regarded as a foundation to build on in future analyses.

[3] In economics, this type of efficiency is known as technical or x-efficiency, as opposed to allocative efficiency.
[4] Bobbie Mills, "Assessing Covid-19 Cost Pressures on England's Schools."
[5] Jon Andrews, "Analysis: 'Levelling up'- What It Really Means for School Funding," Education Policy Institute, accessed September 29, 2021, https://epi.org.uk/publications-and-research/analysis-the-prime-ministers-promise-to-level-up-school-funding/.

Existing methods for estimating efficiency: DEA, SFA and DfE efficiency metric tool

This paper uses a method which is well-developed in academic literature to study *technical efficiency* in schools.[6] This method is Data Envelopment Analysis (DEA). DEA works by estimating the *production possibility frontier* which, loosely-defined, represents the maximum possible output that can be realistically produced given a certain level of input (or alternatively the minimum input required to produce a certain level of output). The production frontier can be estimated by analysing a sample of *decision-making units* (DMUs). DMUs can be anything from firms, factories, hospitals to regions or countries – or schools or school groups. DEA analyses the "*inputs"* and "*output(s)"* of DMUs in order to estimate the production frontier. This is illustrated in Figure 2.1.

Figure 2.1: Illustration of a production possibility frontier with one input and one output.



[6] Geraint Johnes, "Chapter 35 - Economic Approaches to School Efficiency," in *The Economics of Education (Second Edition)*, ed. Steve Bradley and Colin Green (Academic Press, 2020), 479–89, https://doi.org/10.1016/B978-0-12-815391-8.00035-5.

The green curve depicts the production possibility frontier which identifies the maximum possible output that is producible with each level of input. Each green point represents an individual *decision-making unit* (DMU) – in this case, schools. Labelled on the figure are two schools that have the same levels of "input" – Schools A and B. While each have the same level of input, School B achieves double the level of output compared with School A. We can calculate the technical efficiency of these schools from their *distance* from the frontier at that level of input. In this hypothetical example, when input is 6, the highest possible output is 500. School A produces half of that possible amount, 250, and so its technical efficiency is 250/500 which is equal to 0.50. We can therefore say that School A is 50 per cent efficient. Meanwhile, School B lies on the frontier and produces the maximum possible output of 500. School B's technical efficiency is therefore equal to 500/500, giving it an efficiency score of 1.00 and we can say that School B is 100% efficient.

We are mindful that the methodology terms "input" and "output" do not sit easily within education. There is not a set of proscribed inputs that make an effective school or trust, and neither is there a single definition for the desired outputs. In spite of this, DEA has been particularly favoured in economic literature as a method for measuring the efficiency of schools and of other public services, precisely because it allows for multiple definitions of an effective school. The term "inputs" can be read as representing the key decisions that school leaders make about how to employ their resources to ensure positive outcomes. In addition, the term "output" refers to the school's effectiveness in delivering positive outcomes for young people – our measures for this will be detailed in section four.

The following sections describe DEA in detail and outline an alternative method known as Stochastic Frontier Analysis (SFA). Whilst we focus on DEA in this working paper, our future papers will also incorporate SFA and therefore it is useful to compare and contrast the two methods here. They each arrive at the production frontier via very different routes, each having various advantages and disadvantages.

Data Envelopment Analysis (DEA)

Data Envelopment Analysis, or DEA, starts from the premise that DMUs operating within the same industry or service do not all seek to maximise their productivity in the same way. This makes it especially suited to analysing efficiency in public services such as education and health, where inputs and outputs are not strongly defined by prices. Instead, DEA allows that DMUs will favour different combinations of inputs and outputs depending on the 'mission' of their operation. An example given in Johnes' review of economic approaches to school efficiency is that one university might attach more weight to the quality of its graduates, whereas another may focus more on widening participation in its admissions policies.[7] Equally, a training hospital might spend more on teaching resources to achieve positive outcomes for patients compared with a private hospital. Within certain constraints (e.g. practical or statutory constraints), each DMU has its own definition of optimum outcomes and the preferred way of achieving it. The benefit of applying this to education is clear in that it recognises that, subject to various constraints, there are multiple definitions of positive educational outcomes and multiple approaches to providing these to young people.

[7]Geraint Johnes, "Chapter 35 - Economic Approaches to School Efficiency," in *The Economics of Education (Second Edition)*, ed. Steve Bradley and Colin Green (Academic Press, 2020), 479–89, https://doi.org/10.1016/B978-0-12-815391-8.00035-5.

A step-by-step demonstration of DEA is included in section 5. The method expresses efficiency as the ratio of weighted sum of outputs to the weighted sum of inputs for each DMU. The DEA identifies a different set of weights for each DMU through solving a linear-programming problem. For each DMU *x*, the DEA finds the optimum set of weights to attach to each input and output, such that DMU *x*'s output/input ratio is as close as possible to one, with the constraint that applying those same weights to each DMU in the set must not produce an output/input ratio exceeding one.

DEA accommodates multiple inputs and multiple outputs, whereas SFA allows only a single output to be defined. Drawbacks of DEA include (a) the fact that any deviation from the frontier is interpreted as inefficiency, without allowing for random differences ('noise') or data error; and (b) the inability to test hypotheses or to understand the contribution of different explanatory variables to the output measure or to inefficiency.

Stochastic Frontier Analysis (SFA)

Stochastic Frontier Analysis, or SFA, was developed within the regression and least squares family of methods which involves estimating a line of best fit through a series of points. The key difference in the aim of the method is not to find the line that best fits the average values of the dataset, but to create an *envelope* that identifies the 'best-practice' values lying on the frontier. That is to say it finds the best-practice points that *envelope* the rest of the DMUs. Earlier iterations of the method achieved this by adjusting the constant in the regression equation, thereby shifting the line of best fit up or down by a constant amount in order to envelope the full data. More sophisticated methods have since been developed which essentially break the error component of the regression equation down to (i) data error and noise and (ii) inefficiency. An assumption must be made by the researcher about the distribution of the inefficiency of the data (e.g. half-normal or truncated-normal, gamma or exponential), which the statistical model will then use to distinguish inefficiency from random error.

An advantage of this method is that it estimates a parameter for each input and environmental variable, including p values, which enables the researcher to explore the contribution made by each input in determining the output and the inefficiency in the data.

The analysis in this paper has focused more heavily on DEA to identify the most efficient schools and groups. In preparation for this paper we explored an SFA true random effects model which assumed a truncated-normal distribution of the inefficiency term. Our purpose was to explore the impacts of various inputs on improving school performance and the influence of socioeconomic deprivation on inefficiency. However in this instance our algorithm did not converge and therefore did not produce viable results. In future papers we will continue to improve our model and present SFA findings.

Department for Education efficiency metric tool

To better enable schools to benchmark their own financial efficiency against other similar schools, the Department for Education (DfE) developed a simple metric tool.[8] Given that this working paper is concerned with developing an optimal measure of school efficiency, it is useful here to outline this existing benchmarking tool, which works in four steps:

[8] https://www.gov.uk/government/publications/schools-financial-efficiency-metric-methodology

- For a given school (School A), identify 50 similar schools based on that school's level of free school meals (FSM) and special educational needs (SEN)
- Calculate a very simple efficiency score for each school in the set of 50. The calculation is the ratio of pupil progress to funding per pupil.
- Rank each of the 50 schools according to their own efficiency scores
- The ranking enables School A to consider how it compares with this set of 50 'neighbour schools'. It may then enquire directly with the leadership of these peers as to how their practices differ and how it might learn from them to improve its efficiency.

The main limitation of this metric for our purposes is that the metric itself does not explicitly model the different behaviours and resource-deployment pursued by school leaders to achieve their efficient outcomes. Instead, it offers an efficiency score based on two measures that either the school has only marginal influence over (funding per pupil) or is focused on improving regardless of their efficiency (pupil progress). The actual changes that a school would need to make in order to improve efficiency cannot be extracted from this efficiency score, but rather must be uncovered through follow-up conversations among school leaders and their identified 'neighbour schools'. This is not a flaw in the tool in itself as it was designed to be used by individual schools to benchmark themselves against their peers. However the gap we hope to plug with our model is to provide an overview of efficiency of the whole system, and also to make explicit the common features of the most efficient schools.

3. How to compare efficiencies of different types of school group

We have just covered the concept of the decision-making unit (DMU) in the previous section. A key decision in designing this analysis is determining what our DMU is in a way that faithfully reflects the realities of different types of school group. If our objective is to understand the efficiency of school groups, should the DMU be at school- or group-level? Do we consider the inputs and outputs of individual schools, or the combined inputs and outputs of the groups they belong to? Should this vary for different types of school group? We have consulted on this with our advisory group during the design process and we encourage ongoing feedback in light of the results reported in this paper.

The advisory group was split in its opinion on whether the DMU should be located at school- or group-level. For some, it was strongly argued that, for MATs, the DMU must be at trust-level, due to the trust being the legal entity. Others held a similar view that, for MATs, allowing individual academies to be DMUs would assume too much autonomy, especially for those in trusts that have consolidated their budgets. The direct obverse was also argued, with some contributors concluding that treating MATs as the overall DMU would assume a level of MAT control that is not the reality in many cases. Within the advisory group we have leaders of MATs that are highly centralised and others where almost all decisions, including curriculum and staffing, are delegated to individual academies. Clearly MATs are a heterogeneous group, with varied schemes of delegation meaning that decisions can be taken at both the school- and group-level depending on the set up of the individual MAT.

As for maintained schools, one contributor argued that it is technically impossible to locate the DMU, and for this reason the methodology risks failing at the first decision. This is due to the *de facto* and *de jure* sharing of responsibilities between maintained schools and the local authorities, which varies from locality to locality. For example, the LA is mostly, but not always, the legal employer of teachers, but generally makes none of the important education decisions linked to staff recruitment or curriculum.

Overall, our choice of DMU must consider the *de facto* realities of where decisions are made and how resources are deployed, as opposed to *de jure* technicalities of legal responsibility. We are therefore satisfied that school-level is the natural DMU for maintained schools. As for MATs, our judgement is limited as we cannot distinguish between them in terms of their schemes of delegation, and with existing national data we cannot know accurately how resources are deployed between schools.

In this report we have opted to treat all MATs as group-level DMUs, and all other group types at school-level (Figure 3.1). We feel this best reflects the overall reality of decision-making and resource deployment in these groups, whilst we recognise that all group types are heterogeneous and not all individual groups perfectly fit this set up.

Figure 3.1: List of group types used in this analysis and DMU-level for each

Group type	Decision Making Unit is… …school-level	…group-level
Single academy trust	✓	
Local authority	✓	
Foundation trust	✓	
Federation	✓	
Diocese	✓	
Diocesan MAT		✓
Starter trust, <1,200 pupils		✓
Established trust, 1,200-5,000 pupils		✓
National trust, 5,000-12,000 pupils		✓
System trust, >12,000 pupils		✓

For all our DEA analyses, we use all schools in England for which we have valid data for the cohort completing Key Stage 4 in 2019 (Figure 3.2). We require each school to have been part of its group for at least two years at the beginning of the 2018/19 academic year.

Figure 3.2: Total number of groups and schools in each type, and the average number of schools per group

Group type	N groups	N schools	Average N schools per group
Single academy trust	494	494	1.0
Local authority	119	511	4.3
Foundation trust	39	49	1.3
Federation	10	13	1.3
Diocese	56	402	7.2
Diocese MAT	78	171	2.2
Starter trust	38	43	1.1
Established trust	245	355	1.4
National trust	65	199	3.1
System trust	16	152	9.5

In 2018/19, there were 3,965 state-funded schools with key stage 4 results.[9] In this working paper, **a substantial proportion of schools are missing from our final dataset (about 40 per cent).** This is due to a combination of:

- Missing data across the multiple datasets that we use to derive our input variables (e.g. school workforce census, academies accounts return, and national pupil database);
- Suppression rules which require the removal of all schools with fewer than ten pupils finishing key stage 4;
- Schools lost when matching datasets together. Thich is generally caused by their current or predecessor Unique Reference Number (URN) being assigned to different years in different datasets. Future reports will improve our matching to avoid losing data in this way;

[9] This includes local-authority schools, academies, free schools, city technology colleges, further education colleges with provision for 14-to-16-year-olds and state-funded special schools. Key stage 4 performance, 2019 (revised), Department for Education.

- Our inclusion criteria which states that schools must have joined their current group at least two years before the beginning of the 2018/19 academic year;
- And finally removing outliers and extreme values from our input and output variables.

We estimate that about 25 per cent of KS4 schools are removed from our final dataset due to missing data, suppression rules and data matching (leaving about 3,000 schools). A further 17 per cent do not meet our criteria of being part of their group for two years before 2018/19 (leaving about 2,500 schools). Finally, about 100 schools are removed that have improbable input or output values. All in all, about 2,400 schools are included in our final sample, which represents about 60 per cent of all schools with KS4 data in 2018/19.

4. Selection of input and output variables

Key in the construction of our efficiency measure is the selection of input and output variables. Again, we consulted with the advisory group of sector experts and school leaders and arrived at a short list of custom-made input and output variables to be used in this analysis. Our main constraints are (a) the availability of national school-level data, (b) the accuracy and reliability of data where it does exist and (c) the need for parsimony as DEA is particularly sensitive to the number of variables included.

Evidently the work that schools do cannot be adequately reflected in a handful of 'inputs' and 'outputs' that are defined by data points. These methods require a certain level of reduction to the essential 'ingredients' a school or trust uses to achieve positive outcomes for young people. As such, our selected inputs do not aim to capture everything that makes a school effective, but rather they aim to capture the key decisions that school leaders can make about how to deploy their resources.

Discussion of output variable

We use a single output variable in this initial paper. **Our selected output is a 'value-added' measure, which adjusts pupil attainment for the level of challenge in the cohort.** We create this by taking each DMU's average Attainment 8 score and dividing it by the attainment that would be predicted by national data for that school or trust, given the characteristics of their pupil intake. This predicted attainment is derived using multi-level modelling of the National Pupil Database 2018/19. Figure 4.1 illustrates how we create our selected output variable.[10]

Figure 4.1: Illustration of how we create our selected output variable

School or Trust – decision making unit (DMU)	Observed Attainment 8	*divided by*	Predicted Attainment 8, contextualised for pupil and school characteristics*	*equals*	Our output measure: 'Value-added': Number of points achieved for every point predicted. Scores above 1 mean the DMU outperforms its prediction. Scores below 1 mean the DMU performs below its prediction.
DMU 1	60		30		= 60 / 30 = **2.0**
DMU 2	45		45		= 45 / 45 = **1.0**
DMU 3	30		60		= 30 / 60 = **0.5**

Note: All attainment data is from academic year 2018/19.
*Predicted Attainment 8 is estimated with a multi-level model using the National Pupil Database linked to the school census 2018/19. The model estimates Attainment 8 for each pupil accounting for their prior attainment, gender, socio-economic deprivation, first language, special educational needs and ethnicity. The model also accounts for school-level characteristics: mean and standard deviation of overall school prior attainment. Pupil estimates are averaged to produce a school-level prediction. Due to data suppression rules we only include schools with at least ten pupils.

[10] Alternative value-added measures such as Progress 8 or contextualised value-added measures which can be constructed using modelling are not well suited to DEA analysis because they are centred around zero.

In real-world terms, this measure represents the extent to which a school or trust enables its pupils to achieve higher than they would on average in any school, based on their characteristics such as prior attainment, gender, and disadvantage. The score represents the number of points achieved for every point that was predicted. Scores above 1 mean the pupils in the DMU achieve higher than their predicted attainment. Scores below 1 mean pupils in the DMU achieve lower attainment than would be predicted.

We choose to employ a single attainment-based output variable because, even though an advantage of DEA is that it allows multiple outputs to be defined, DEA is subject to the assumption of *free disposability.* This assumption states that all inputs and outputs are *freely disposable* and that any individual input or output is assumed to be non-essential to successful production and could be acceptably disposed of in the name of efficiency. Therefore, the basic DEA model is not set up to reward the DMU that achieves a good *balance* of all the outputs specified, but instead the DEA will often identify which *particular* output the DMU is strongest at producing and will 'zero-weight' the rest. The effect of this is that the DEA indicates that the DMU should focus more of its resource on producing its strongest output and disregarding the other outputs it is less strong in. The DEA is therefore prone to rewarding a *trade-off* between two outputs as opposed to finding a *balance* between both.

Our judgement is that, from a standpoint of national policymaking for compulsory education, there are few outcomes that could be acceptably traded off against one another. For example, whilst we do not wish to reward schools that achieve high attainment at the cost of pupil wellbeing, neither would we wish to reward schools with excellent pupil wellbeing that do not enable their pupils to achieve a basic level of attainment. The same could be said for pupil destinations, attendance, and so forth. Our decision is to focus on a single attainment-based output in this initial paper, with the intention of exploring multiple outputs, including 'negative' outputs such as high teacher turnover, and models which relax the assumption of free disposability in future papers.

Discussion of input variables

We are grateful to the advisory group for contributing a detailed discussion of which input variables are most important to include to capture how school leaders make use of resources to impact on pupil outcomes. Here we summarise the range of areas highlighted in the discussion, many of which could not be included in this model due to lack of adequate or reliable national data.

Staffing was a major theme in the discussion of input variables. The key decisions on staffing highlighted by the advisory group were:

On teaching staff:

- Decisions related to the age and experience of teaching staff, which are also related to pay and progression. Some contributors pointed out that employing young and inexperienced staff can be a cheap operating model, but that this can be unsustainable due to burnout and turnover
- Pupil teacher ratios (PTR), which can be driven by curriculum-led financial planning (CLFP)
- Contact ratio, and planning, preparation, and assessment time (PPA)
- Investment in continuing professional development (CPD) and leadership development
- Talent management and staff evaluation – including managing out those who underperform

- Staff pay, which depends on whether the school adheres to national pay scales

On leadership teams:

- Size and structure of leadership team
- Number of headteachers or executive headteachers across multiple schools

On support staff:

- Number and type of support staff
- Support staff salary, related to quality and experience

Overall, in a group of multiple schools, particularly a MAT, decisions must be made regarding the degree of control over staffing structures and recruitment of staff of various seniority levels, and whether these decisions are centralised or dispersed among the group.

Decisions on **curriculum** provision were also a common theme. These include:

- The overall shape and breadth of curriculum, in response to local need. One contributor commented that the constraints of the National Curriculum, headline measures and expectations of Ofsted meant that most schools end up having similar staffing structures to deliver the core curriculum
- Spending on educational resources, including how much central control exists at trust-level over this spending
- Degree of alignment across multiple schools in terms of curriculum and assessment points
- Extracurricular provision, character education and school ethos
- Structure of the school day
- Progression across age sectors
- Long-term investment of funding for capital contingency and curriculum

Key decisions on **budgeting** include:

- In a MAT, a key decision is whether and how to consolidate funding across multiple schools in order to set school budgets based on school need as the group sees fit
- As noted above, decisions must be made regarding the degree of central control over certain spending areas such as staffing, resourcing and CPD
- Value for money in service procurement: reducing costs across bought-in services to free up funding to spend on teaching and learning
- Long-term investment for capital, contingency, and curriculum
- Decisions on building maintenance costs
- Creating self-generated funding within the school or trust
- One contributor commented that running a post-16 can be a drain on resources

Pastoral support and pupil inclusion were also highlighted as key decisions for use of resources that can impact on pupil outcomes. These include:

- Policies on admissions, behaviour and exclusion. The combined effect of these can influence the characteristics of the pupil cohort
- Managing relationships with families and carers
- Use of pupil premium

- Broader support services – especially for those with additional learning needs – including pastoral support, counselling, and breakfast provision. One contributor warned that in all schools there will be high-needs individuals, however a school “could decide not to meet those needs”, which could enable it to look more efficient without a substantial impact on overall pupil outcomes

These lists include both positive and questionable practices and decisions that education centres could pursue. Nevertheless, they are all *decisions* related to resource-provision that rest entirely or partially within the control of school leaders which can have an influence over pupil outcomes.

From this list we have considered what reliable national data is available, and which are most important to include in a cross-sectional assessment of efficiency in academic year 2018/19, as opposed to a longitudinal study that we may produce in future reports. Our final selected inputs and output are set out below.

A key blind spot is that we have no comprehensive view over how much central control is exercised within each school group, and this is a particular yoke for identifying how funding is shared between and within schools. Another blind spot is a lack of data on centralised staff, particularly school improvement teams within MATs, as these roles are not covered by the School Workforce Census.

Final selected inputs and output

Following consultation as outlined above, our selected inputs and outputs are set out in Figure 4.2, including details on sources and how each was constructed. All in all, we have four input variables: two covering staffing; and two covering spending. As mentioned above, these are necessarily a reductive view of a school’s operations, and they aim to capture the key decisions that leaders make about how to deploy their resources. We welcome feedback through the consultation which accompanies this paper’s publication.

For staffing measures, we have selected (1) **the size of the leadership team**, adjusted for pupil full-time equivalent (FTE), and (2) **the total number of years of qualified teaching experience employed within the school or trust**, weighted by the FTE of experienced qualified teachers and adjusted for pupil FTE. By ‘experienced qualified teachers’ we mean those with at least five years of experience since qualifying. Whilst national data on school spending on teacher salary is available, this is not appropriate to use in efficiency analysis because prices for teachers vary nationally, therefore spending in one part of the country is not equivalent to spending in another. We instead choose years of qualified experience as an indicator of teacher quality, though we recognise this is a limited indicator. A drawback of our leadership variable is that it does not indicate quality. In future reports it would be possible to construct a similar measure that shows number of years of experience among leaders.

On spending variables, for the same reasons as set out above related to prices varying across the country, our spending variables are given as percentages of overall expenditure as opposed to amount spent in GBP. We have not included spending on staff development, because we deem the available data on this to be too inconsistent. Future reports may include spending data on educational resources, and future longitudinal analysis could also include long-term capital investment.

Figure 4.2: Summary of selected input and output variables, including details of source and construction

	All KS4 state-funded schools with complete data in 2019	Source and details of construction
	Output	
	‘Value added’ - Average **Attainment 8** score (2019) **divided by** **predicted Attainment 8** of pupil cohort, contextualised for pupil and school characteristics. See Figure 4.1 for further explanation.	Attainment 8: National Pupil Database Predicted Attainment 8: Estimated with a multi-level model using the National Pupil Database linked to the school census. The model estimates Attainment 8 for each pupil accounting for their prior attainment, gender, socio-economic deprivation, first language, special educational needs and ethnicity. The model also accounts for school-level characteristics: mean and standard deviation of overall prior attainment. Pupil estimates are averaged to produce a school-level prediction. Due to data suppression rules we only include schools with at least ten pupils.

Figure 4.2 continued

	All KS4 state-funded schools with complete data in 2019	Source and details of construction	Averaged for each year the 2018/19 cohort was in the school
	Inputs		
Staffing	FTE of **Leaders**, per pupil FTE	Publicly available school-level timeseries derived from School Workforce Census, Teacher and support staff Full-time Equivalent and headcount numbers for schools, DfE.[11] Combined with publicly available school-level timeseries of pupil FTE (schools, pupils and their characteristics).[12]	✓
Staffing	Total years of **qualified teaching experience** in the DMU, per pupil FTE, weighted by FTE of experienced qualified teachers	Constructed from School Workforce Census contract data. Years of experience are calculated as the number of days to the closest year between the census date and the date of qualification. This total is then weighted by the FTE of experienced qualified teachers. We defined 'experienced' teachers as those with at least five years' experience since qualifying. We weight the total years because more experienced teachers are more likely to have lower FTE, and are also most likely to have a positive impact on pupil outcomes. Due to suppression rules we only include schools with at least five classroom teachers	✓
Spending	% of total expenditure spent on **education support staff**	Academies Account Return and Consistent Financial Reporting datasets.	✓
Spending	% of total expenditure spent on **back office functions**[13]	Academies Account Return and Consistent Financial Reporting datasets. Back office expenditure comprises spending on administrative and clerical staff, administrative supplies and bought-in professional services.	✓

[11] School workforce in England, https://explore-education-statistics.service.gov.uk/find-statistics/school-workforce-in-england#dataDownloads-1

[12] Schools pupils and their characteristics, https://explore-education-statistics.service.gov.uk/find-statistics/school-pupils-and-their-characteristics#dataDownloads-1

[13] Initially we proposed to use the reverse of the back office spending variable i.e. the percentage of school expenditure that is *not* spent on back office, as logically more money saved on back office expenses means more money freed to spend on teaching and learning. However the nature of the DEA analysis means that lower input values are rewarded over larger ones. Therefore a school spending a tiny amount *not* on back office would be counted as more efficient.

In theory, our selected inputs should have a positive relationship with the output, whereby more of an input should lead to more of the output. A simple correlation matrix of our full menu of possible inputs and outputs (full table in Appendix) shows that our inputs and outputs do not correlate in a straightforward way. For example, spending on education support staff is quite strongly negatively correlated with school attainment. This is not because teaching assistants are detrimental to school attainment, but rather that schools with lower attaining cohorts make more use of teaching assistants than high performing schools in less challenging contexts. This single example demonstrates that we should not base our decisions solely on correlation – as these figures are far from showing us the causal direction of relationships.

Our selection of input variables is therefore also based more on knowledge and theory of what should reasonably improve outcomes in schools.

For each input variable, we take an average of each DMU's value across all the years the cohort were educated in that school. This is to reflect the overall experience of attending the school from Year 7 to Year 11, rather than taking a snapshot of input levels just before the students sat their exams. We assume the whole cohort attended the school for the full period and we do not allow for mobility between schools.

For all our MAT DMUs, their inputs and outputs are created at trust-level. For example, to create the trust-level output variable we take the total Attainment 8 scores of all schools within the trust and divide it by the total number of pupils finishing KS4 in that trust. Equally, for the spending variables we take the total trust-level spend on, say, education support staff, and divide this by the total expenditure across the whole trust.

Variation of staffing, spending and attainment between different types of school group

Here we examine the variation between different types of school group along some of our selected inputs and output. The boxplots presented in this section demonstrate that in some cases there are clear differences between types of school group and how they deploy their resources.

For example, Figures 4.3 and 4.4 demonstrate the distribution of our selected staffing variables across different types of school group. We find that teachers in multi-academy trusts tend to have fewer years' experience of teaching since qualifying, in comparison with other types of group. System leader trusts and Diocesan MATs appear to have the lowest levels of qualified teaching experience. It can be read from these charts that teachers in system leader trusts and Diocesan MATs have, on average, a little under half-a-year's qualified teaching experience for every pupil in the trust, or about five years' experience for every ten pupils in the trust.

Meanwhile the distribution is more even across groups in terms of size of leadership team. We find that Federations tend to have the largest leadership teams relative to their pupil numbers: the median for Federations is just under one leader (full FTE) for every 100 pupils.

Figure 4.3: Total years of qualified teaching experience in the DMU, per pupil FTE, weighted by FTE of experienced qualified teachers, by group type

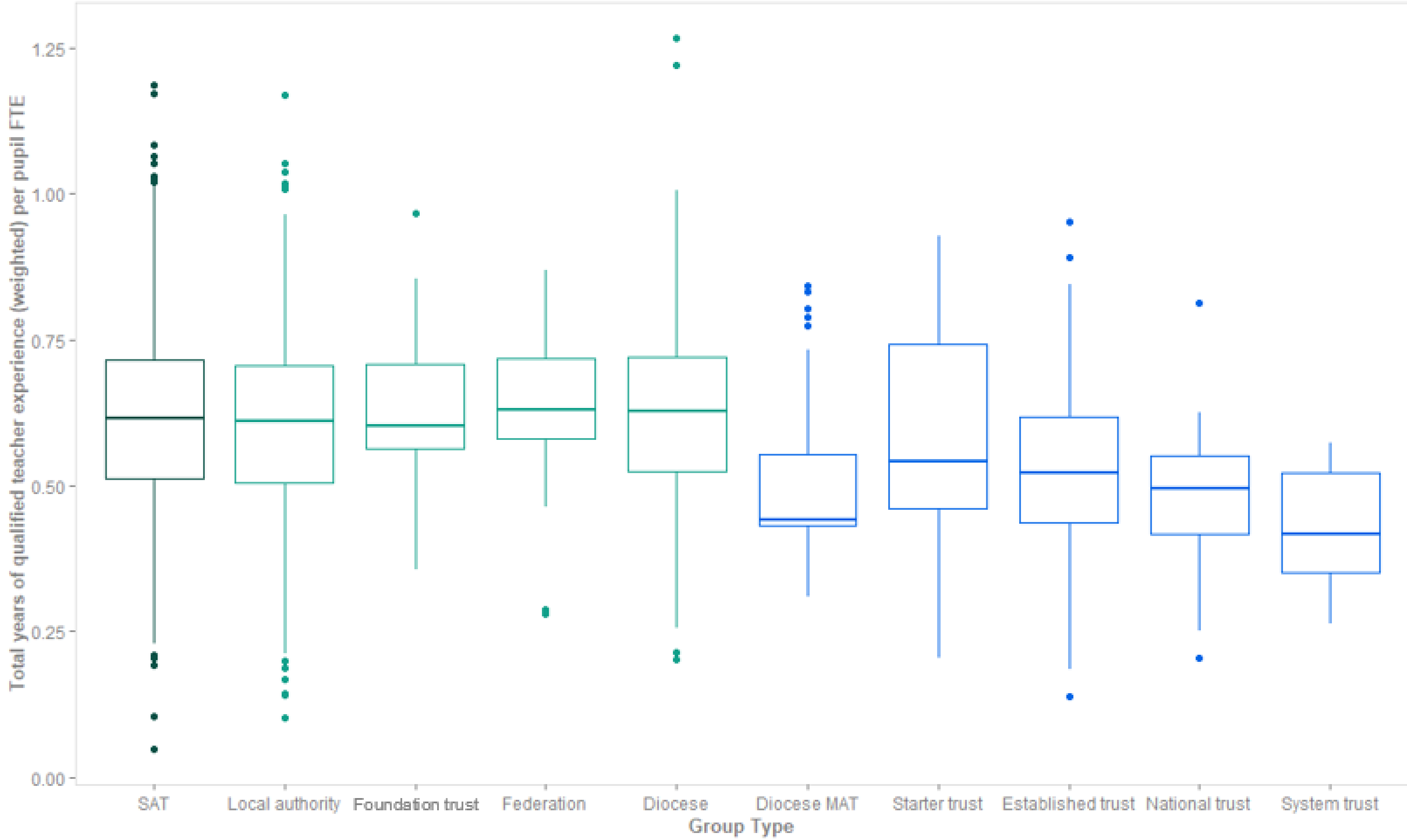


Figure 4.4: School leaders full-time equivalent adjusted for pupil FTE, by group type

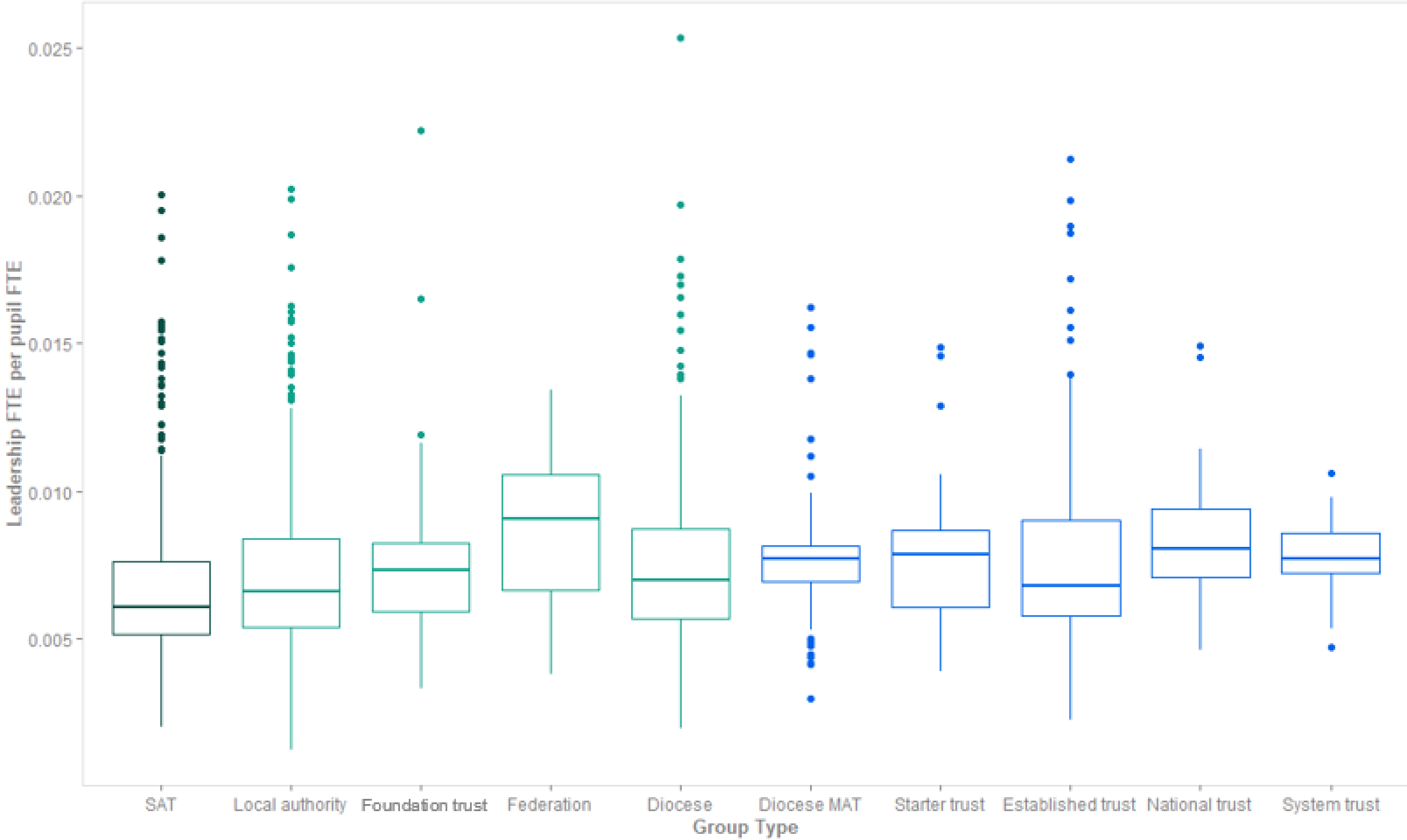


Figure 4.5: Spend on 'back office' functions, as percentage of total expenditure, by group type

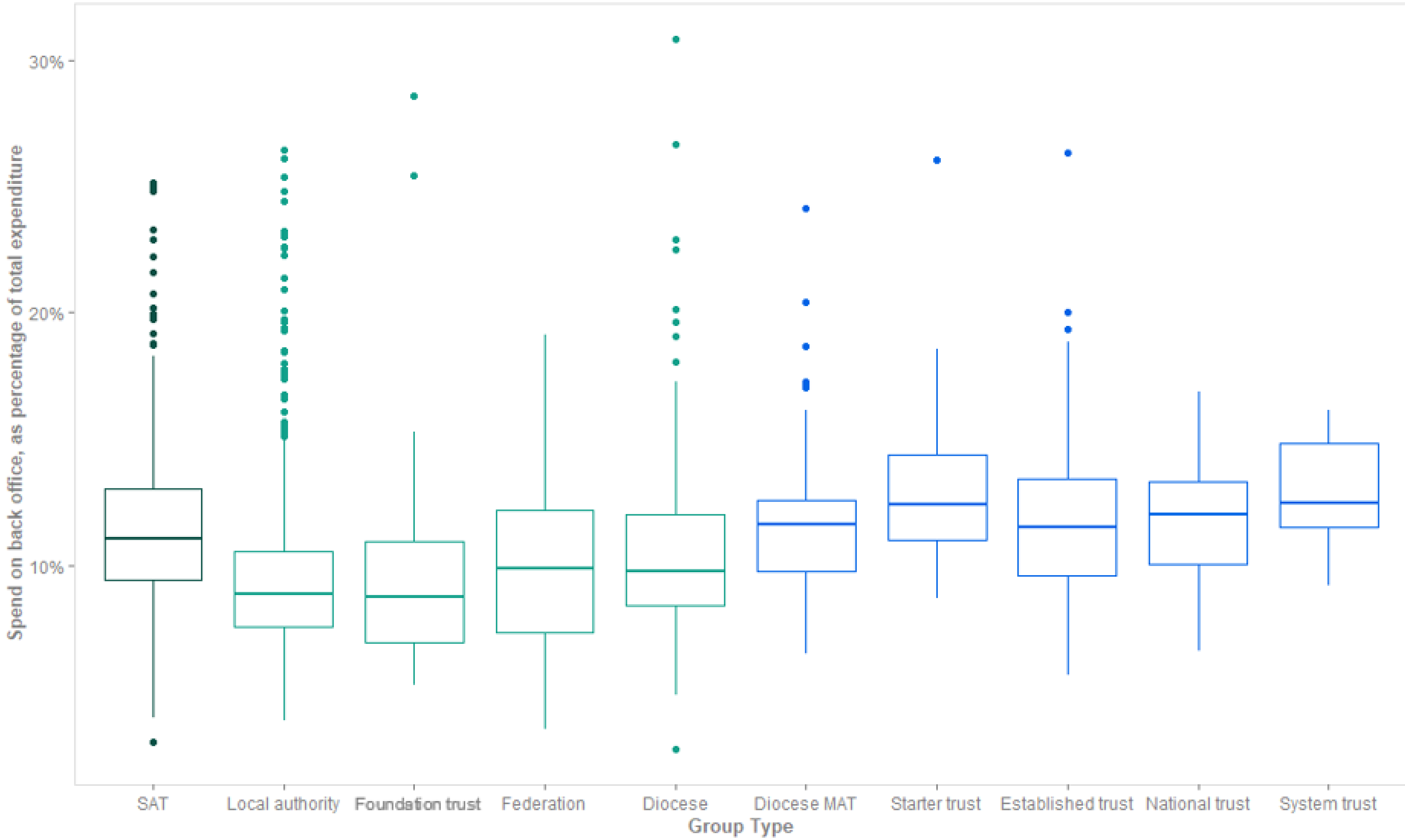
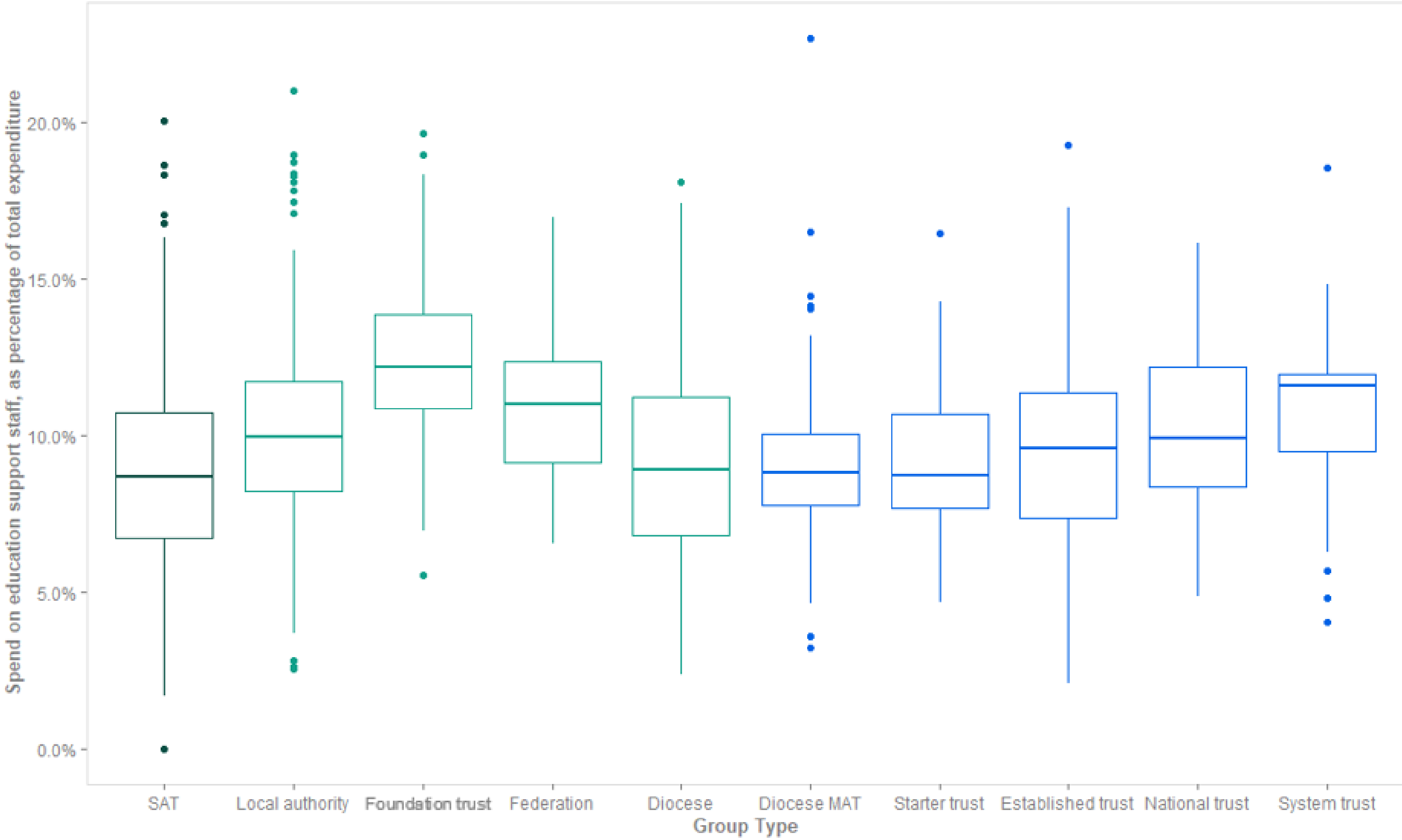


Figure 4.6: Spend on education support staff, as percentage of total expenditure, by group type



Figures 4.5 and 4.6 lay out spending behaviour across different types of group. We find that MATs, and particularly larger MATs, have slightly higher spends on 'back office' functions as a proportion of their total expenditure.[14] We also find that larger MATs have comparatively higher spends on education support staff, and this is likely connected to the fact that larger MATs tend to take on more challenging schools where education support staff may be made more use of. Similarly, we also find that foundation trusts have particularly high spends on education support staff.

Finally, Figures 4.7 and 4.8 illustrate the distributions of our value-added variable (Attainment 8 divided by predicted attainment), and secondly the predicted attainment of cohorts attending different types of school group. We again find that MATs have increasingly higher value-added scores as their size increases. Some single-academy trusts (SATs) and local authority schools have very high value-added, with others achieving very low value-added scores. This can be read in comparison with the predicted attainment set out in Figure 4.8, which shows that many SATs have extremely high predicted attainment based on the characteristics of their pupils, as do some LA schools, Diocese schools, and established trusts. Meanwhile foundation trusts and system trusts both tend to have pupil intakes with very low predicted attainment, however we know from Figure 4.7 that system trusts contrast with foundation trusts in that they achieve some of the highest value-added (actual Attainment 8 divided by their predicted attainment).

Overall, we see substantial crossover between group types and there is more variation within group types than between them. This notwithstanding, there are some observable differences which may bear out as we compare efficiencies between groups.

[14] This finding echoes previous EPI analysis. "Understanding School Revenue Expenditure | Part 6: How Much Are Schools Spending on 'Back-Office' Functions?," Education Policy Institute, accessed September 29, 2021, https://epi.org.uk/publications-and-research/back-office/.

Figure 4.7: Value-added - average Attainment 8 divided by predicted attainment, by group type

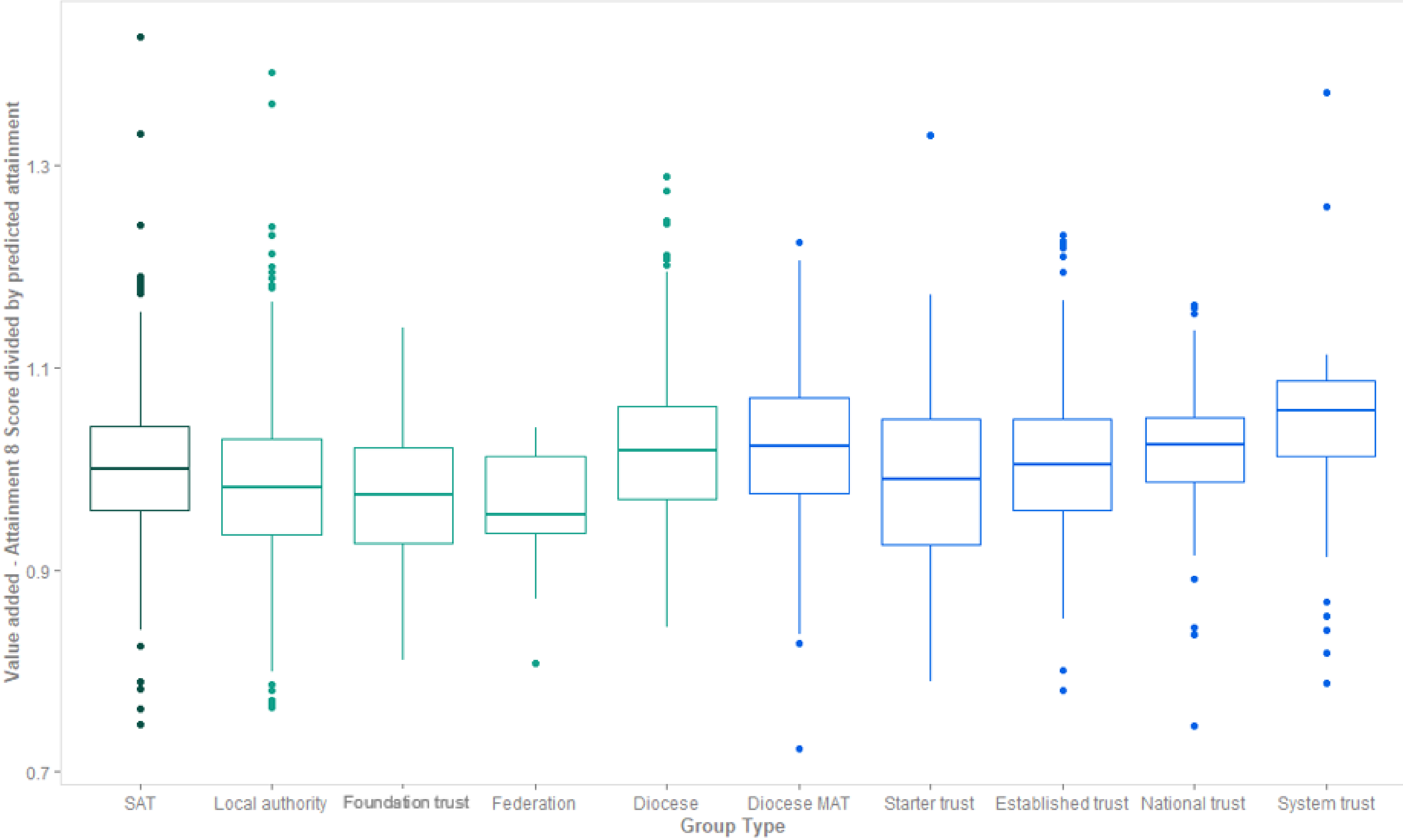
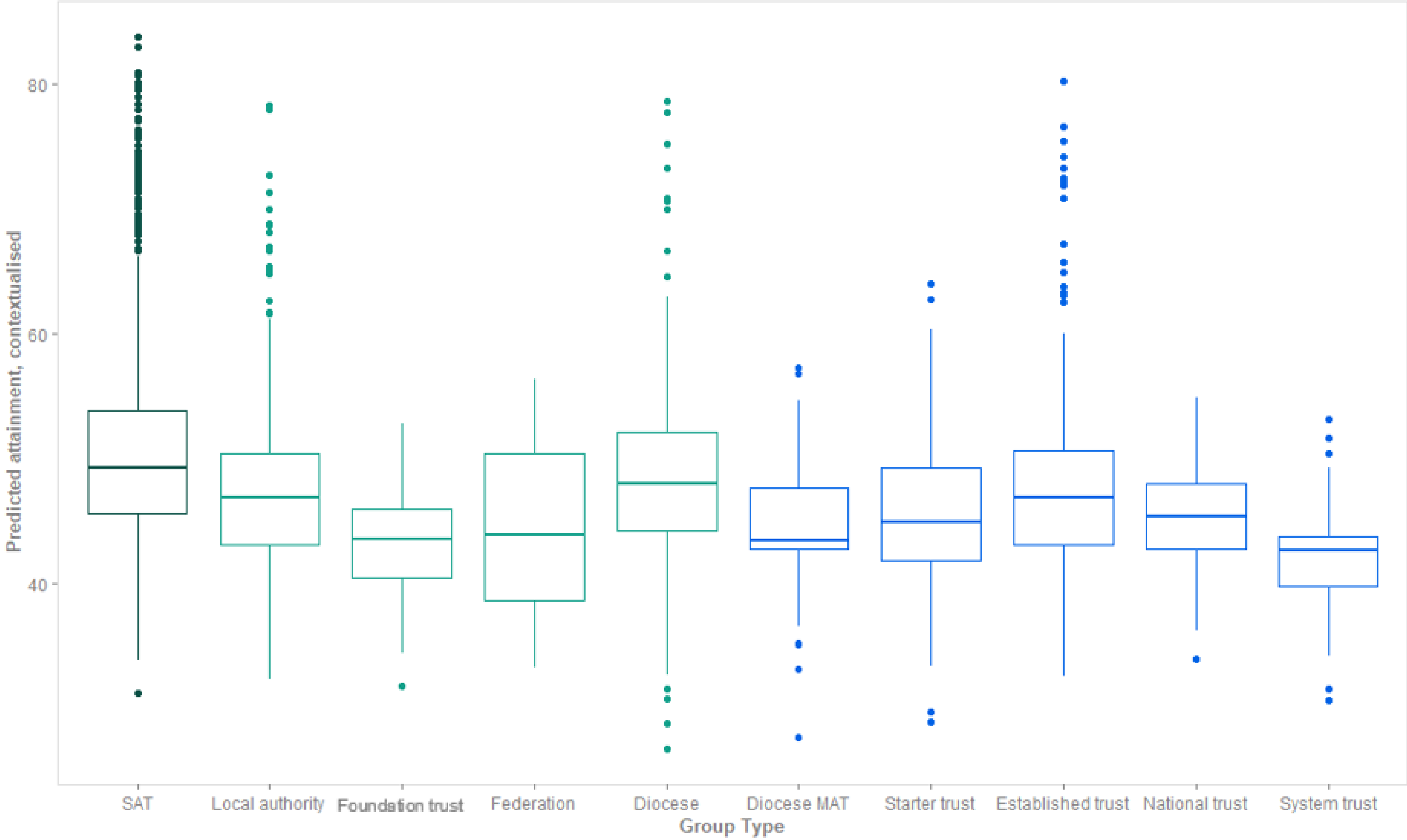


Figure 4.8: Predicted Attainment 8, contextualised for cohort characteristics, by group type



5. Demonstration of DEA: Two input and one output example

Our full DEA model will include four inputs and one output. Prior to this, we present in this section a simplified two input example. This is because, as more inputs are added to a DEA model, the results become more difficult to visualise, and to interpret instantly and intuitively. Our output measure in this example is the same 'value-added' as we will use in our full model. The input measures we select in this two-input example are:

- The full-time equivalent **(FTE) of school leaders**, divided by pupil FTE. School leaders include executive headteachers, headteachers, assistant, deputy heads and other equivalent roles.
- The **FTE of qualified experienced teachers**, divided by pupil FTE. We define a teacher as 'experienced' if they have been qualified for at least five years. **Note** that that this is a different teacher variable to the one we use in the full model. The full model will use total years' qualified teaching experience (per pupil FTE). We do this because the teacher FTE variable makes for easier comparison with the leader FTE variable and therefore lends itself better to this demonstration exercise.

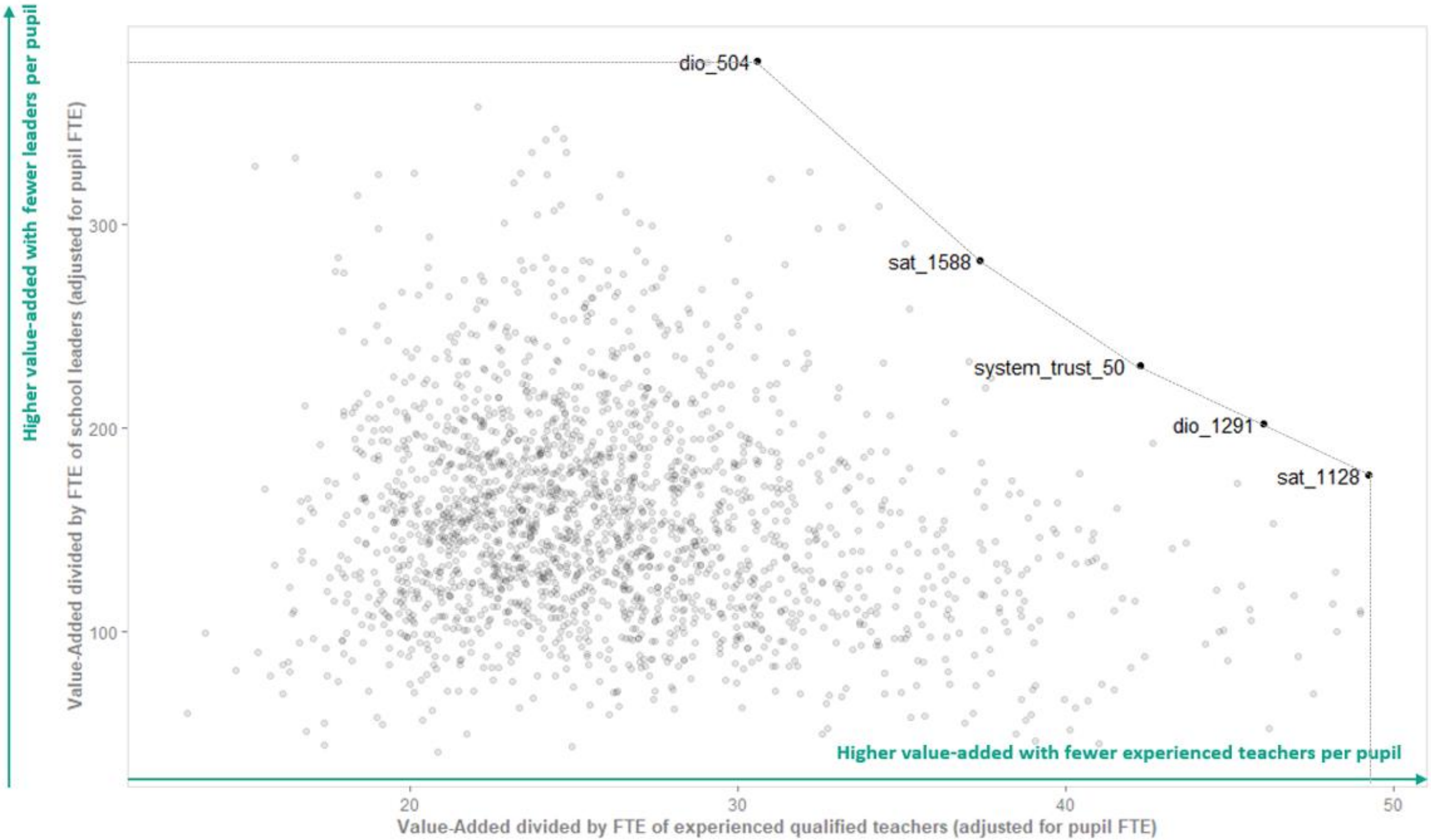
DEA analysis – two-input example[15]

In a two-input DEA, we can use a simple scatter chart to visualise the *production possibility frontier* which was outlined in the introduction of this report. To do this, for each DMU we *divide its output by each of its input values.* These *output-input ratios* effectively tell us how much output a DMU is achieving per unit of each of its inputs and are plotted on Figure 5.1. For example, the DMU we have labelled 'system_trust_50' has five leaders and 26 experienced qualified teachers for every 1,000 pupils (so its input values are 0.005 and 0.026 respectively). Its value-added score is 1.09, indicating that with this level of staffing, pupils in the trust, on average, achieve 1.09 attainment points for every point of predicted attainment. Dividing output by each input, this DMU's output-input ratios are 218 and 42 respectively.

The higher the ratio, the more output has been achieved for each unit of input, and therefore the higher the efficiency. Figure 5.1 labels the fully efficient DMUs which lie on the production possibility frontier. The dashed line joining these DMUs represents that frontier and demonstrates how the points 'envelope' the other less efficient DMUs. We say these schools on the production frontier are 100% efficient or that they have an efficiency score of 1.00. Schools that do not lie on the production frontier have an efficiency score between 0 and 0.99. This score is calculated as the ratio of (a) the distance between a DMU and the axis origin to (b) the distance between the origin and the frontier when drawing a straight line from the origin through the DMU to the frontier.

[15] The DEA model reported here is output-oriented and assumes constant returns to scale. We treat all academy trusts as single *decision-making units* (DMUs), and as such we take their input and output variables at the level of the entire trust. For non-academised groups (e.g. local authorities, federations, and diocese schools that do not operate as academy trusts), each individual school is treated as its own DMU. Note that we remove extreme outlying variables from our inputs and outputs as DEA is particularly sensitive to these cases.

Figure 5.1: Two-input example: Efficiency of KS4 schools and MATs



The DMU we have labelled 'sat_1128' is an example of a SAT that is assessed to be fully efficient due to having a particularly high ratio of output to number of experienced teachers (49.22). Sat_1128 has 25 experienced qualified teachers for every 1,000 pupils, and its pupils achieve 1.24 Attainment 8 points for every point predicted based on their characteristics. This makes for a ratio of about 49 (1.24 divided by 0.025), or more intuitive would be to divide 1.24 by 25 and say that for every thousand pupils, each teacher enables them to achieve 0.049 points higher in Attainment 8 for every point they were predicted.

By contrast, dio_504 has one of the lowest output-to-teacher ratios (30.62), and yet it is still efficient and located on the production possibility frontier. This is because it achieves high value-added (1.14) with a particularly low number of leaders per pupil (three leaders for every thousand pupils). This demonstrates that, within this methodology, it is recognised that there are multiple ways of successfully running a school: two DMUs can have very different combinations of inputs and yet they can each be assessed as equally efficient.

Not labelled on Figure 5.1 are the DMUs with lower efficiency scores (i.e. those that do not lie on the production possibility frontier). We know from the full set of results (not published in this working paper to avoid identifying individual school groups), that some of the least efficient DMUs have the same value-added score (1.09) as two of the fully efficient DMUs. These schools are ranked as less efficient in this two-input example because they achieve a value-added score 1.09 with significantly higher staff inputs. For example, one less efficient DMU has 16 leaders and 68 experienced qualified teachers per thousand pupils. According to this DEA analysis, we know it is *possible* for a DMU to achieve the same level of output with substantially lower staff inputs, and therefore DMUs with higher staff inputs are classed as running with inefficiencies.

This highlights an important feature of DEA: it is assumed that a lower level of input is always desirable. A simple DEA like the two-input example presented here does not take into consideration sustainability, or possible risks of operating with a reduced workforce due to staff burnout and lower retention. We are mindful of these drawbacks and, as we develop our models, we wish to include indicators of sustainability.

6. Full DEA model: Four inputs, one output

In this section we present and discuss results from our full DEA model which uses four inputs and one output.

As outline above, the full set of inputs in our model are:

- **Teacher experience:** the total combined years of experience among qualified teachers in the DMU, weighted by the FTE of qualified teachers with more than five years' experience, and divided by pupil FTE. (Note this is a broader measure than the teacher input we used in the two-input example).
- **Leadership FTE:** The FTE of leadership teachers in the DMU divided by pupil FTE. Leadership teachers include executive headteachers, headteachers, deputy and assistant headteachers, advisory teachers and those with equivalent pay ranges.
- **Expenditure on education support staff:** the percentage of DMU expenditure that is spent on education support staff. We prefer this over the teaching assistant variable because spending on education support staff captures a wider category.
- **Expenditure on 'back office' functions:** the percentage of DMU expenditure that is spent on 'back office' functions.

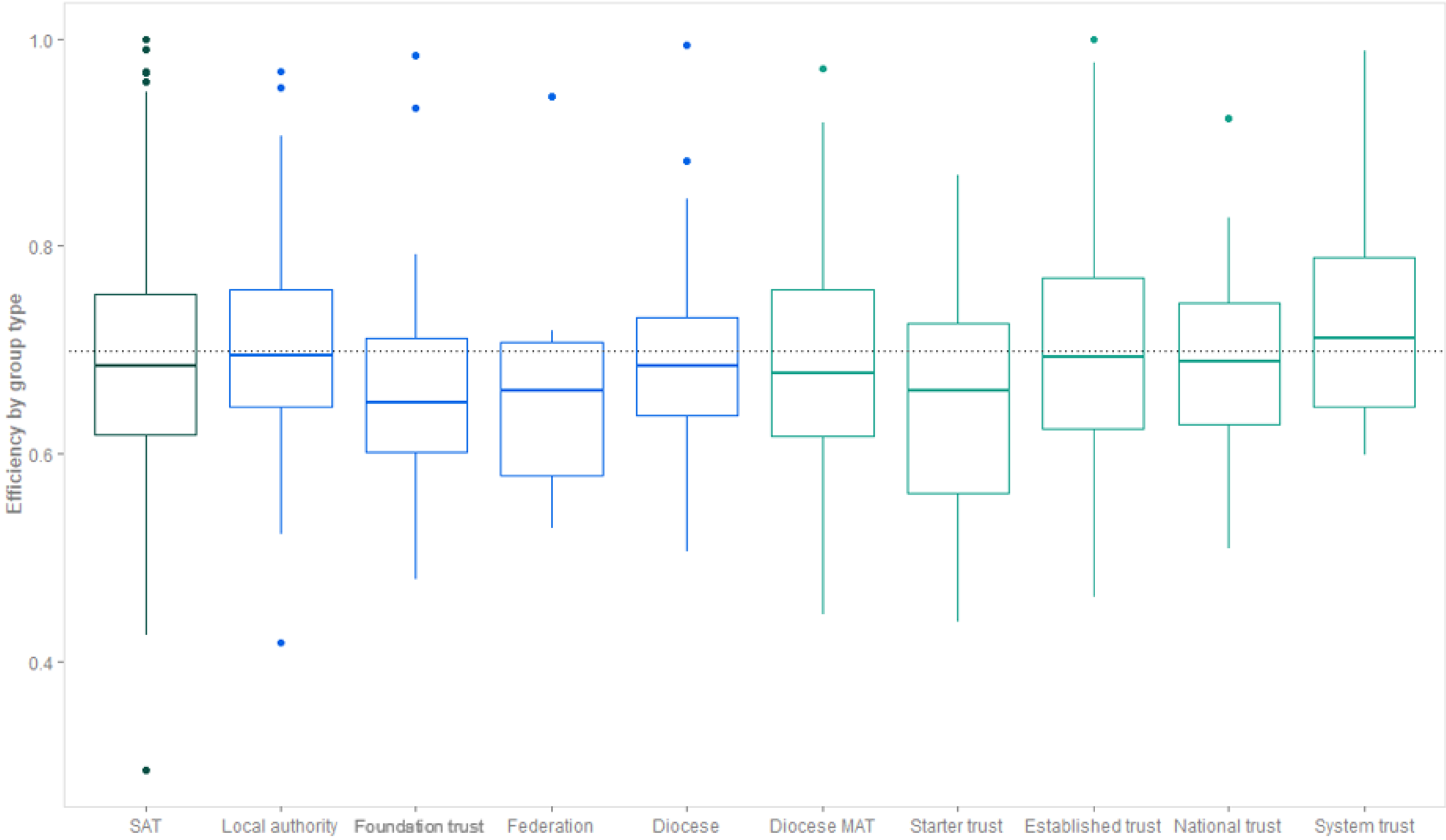
DEA analysis – full model, four inputs[16]

Using all four inputs, we find the overall efficiency in the system among secondary schools in England in 2019 is 0.698. We summarise the overall findings in Figure 6.1, which uses boxplots to illustrate the spread of efficiencies among our different group types. The dotted line represents the overall mean efficiency in the system. The table below summarises efficiencies by group.

Group type	Median efficiencies by group type	N Efficient DMUs in group type
Single academy trust	0.685	11
Local authority	0.695	8
Foundation trust	0.649	0
Federation	0.661	0
Diocese	0.685	7
Diocese MAT	0.677	0
Starter trust	0.661	0
Established trust	0.694	4
National trust	0.689	0
System trust	0.712	0

[16] This model follows the same specification as the two-input example presented earlier (output-oriented and assuming constant returns to scale), using the broader set of inputs as discussed. As previously, our *decision-making units* (DMUs) vary depending on the type of group: academy trusts are treated as single trust-level DMUs, whereas non-academised groups are split into individual school-level DMUs. We have removed extreme and outlying cases from our datasets as DEA is particularly sensitive to these.

Figure 6.1: Full DEA analysis, efficiencies of KS4 schools and MATs by group type



Most group types have a median efficiency very close to the overall average efficiency, ranging between 0.66 and 0.69.[17] System leader trusts have the highest median efficiency of 0.71, followed closely by local authority schools (LAs) and established trusts with median efficiencies of 0.69. Foundation trusts have a substantially lower median efficiency of 0.65, and the boxplots also show that nearly 75 per cent of federations and foundation trusts have an efficiency below the overall average.

The picture is slightly different if we consider which group types have the most fully efficient DMUs (efficient DMUs have an efficiency score of 1.00). As reported in the table accompanying Figure 6.1, eleven single-academy trusts are fully efficient under this model, as are eight LA schools, seven Diocese schools and four established trusts. Otherwise, all other group types have zero fully efficient DMUs.

It is notable that only four MATs out of about 450 come out as fully efficient in this model. We interpret this as being due to an idiosyncrasy of our model brought about by attempting to treat MATs as unified groups and others as individual schools: For all our MAT DMUs, their inputs and outputs are created at trust-level. For example, to create the trust-level output variable, we take the total Attainment 8 scores of all schools within the trust and divide it by the total number of pupils finishing KS4 in that trust. Equally, for the spending variables we take the total trust-level spend on, say, education support staff, and divide this by the total expenditure across the whole trust.

The intention is to reflect the way resources are deployed across a MAT as a whole, as opposed to treating each academy as a standalone unit. However, a consequence of creating a single measure across multiple schools (which each might have very different levels of inputs and outputs) is that larger or smaller values are “smoothed” out. Trust-level variables therefore “regress towards the mean”. The outcome of this is that, because trust-level inputs and outputs are less likely to be either particularly high or low, the DEA is less likely to locate them on the frontier. This is especially likely to be the case for larger MATs with larger numbers of schools, as values are progressively more “smoothed’” out by summarising across a larger number of units.

Meanwhile schools that are part of single academy trusts, local authorities or non-academised Dioceses with particularly high or low levels of input and output are left unadjusted, and therefore disproportionately more likely to lie on the frontier. This is therefore a limitation of this model as it stands: **this DEA model is artificially less likely to find MATs efficient due to their values being summarised across them and thereby regressing towards the mean.**

Nevertheless, whilst MATs appear less likely to be classed as fully efficient, they still have comparatively high overall efficiencies as a group, with system trusts having the highest median efficiency.

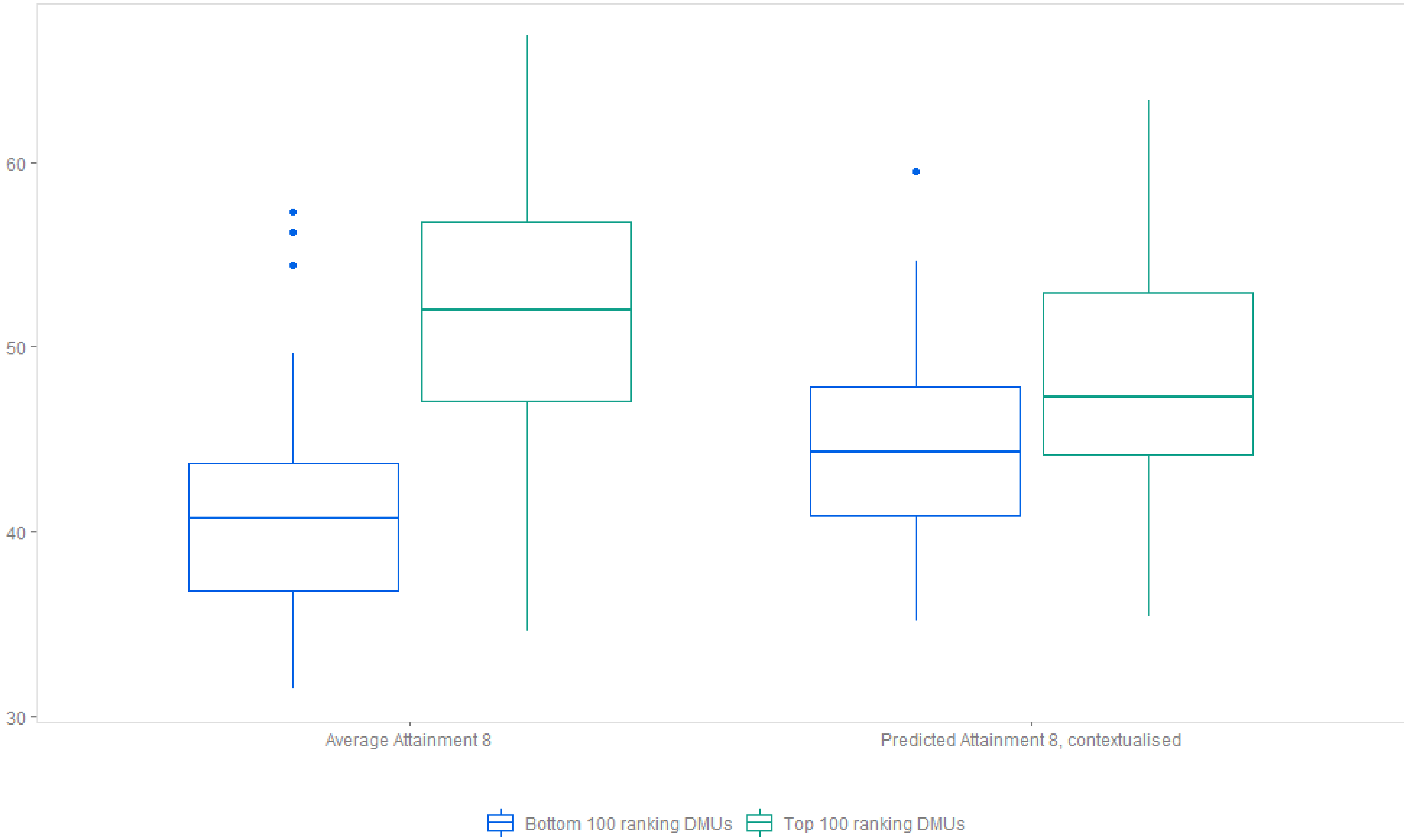
[17] To find median efficiencies by group type, we first find the geometric mean of efficiencies of individual groups with multiple DMUs. For example, for a specific Local Authority, we find the geometric mean of the efficiency scores of all schools (DMUs) in that Local Authority. This gives a single efficiency score for each individual group. We then take the median of these by group type.

Comparing the features of the most and least efficient schools and trusts

Turning now to consider the DMUs that rank as the most and least efficient according to this model, we can compare their inputs and outputs to explore their characteristics. Figures 6.2 to 6.4 compare input and output values of the top and bottom 100 DMUs ranked by efficiency score.

We find that the most efficient DMUs under this model have a range of predicted attainment (Figure 6.2). The schools and trusts ranking within the top 100 for efficiency have a fairly overlapping range of predicted attainment to those in the bottom 100, though overall the most efficient do have higher predicted attainment on average. This demonstrates that some schools and trusts with particularly challenging cohorts are achieving high value-added (their actual Attainment 8 is higher than would be predicted given pupil characteristics) with comparatively low levels of input. This is evidence that our DEA model is capturing efficiency, as opposed to picking out those schools with high performance profiles in terms of overall attainment.

Figure 6.2: Full DEA results: Comparison of inputs and outputs between the top and bottom ranking DMUs. Attainment variables.



Comparing the **staffing decisions** between the highest and lowest ranking 100 schools and trusts in our full DEA model (Figure 6.3), we find that the most and least efficient DMUs have somewhat similar distributions in terms of the level of teacher experience they make use of. Whilst the median years' experience of qualified teaching among the most efficient DMUs is lower at 0.5 per full-time pupil in comparison with the least efficient DMUs (about 0.6 years of qualified teaching experience per full-time pupil), we find the upper and lower quartile ranges of the two overlap. This suggests that, while the most efficient DMUs might make use of slightly less experienced teachers, high efficiency does not tend to be achieved with *significantly* less experienced teaching staff. The schools and trusts that are classed as most efficient under this model appear to have chosen to use only slightly less experienced teaching staff in comparison to other less efficient schools. We see from the boxplots of the other input variables that greater efficiencies are more commonly made elsewhere.

For example, the size of leadership teams in the most efficient DMUs is notably lower than in the least efficient DMUs. The median of this variable is about five leaders per 1,000 pupils among the most efficient schools and trusts, compared with ten leaders per 1,000 pupils among the least efficient schools and trusts. Overall, this suggests that larger efficiencies are more commonly found in terms of the size of the leadership team relative to pupil numbers than in using less experienced teaching staff. Our leadership variable does not capture quality or experience of leaders, and it is possible that the most efficient schools favour a smaller team of more experienced leaders. Alternatively, it may suggest that top-heavy staffing structures are less effective at realising school improvement. Equally it is likely that the size of the leadership team is not directly related to the value-added achieved by the school or trust, but rather the size of the leadership team is related to different types of groups that have particular performance profiles. For example, we know from Figures 4.4 and 4.7 that Federations have the highest levels of leaders per pupil and some of the lowest value-added scores.

Turning to Figure 6.4 which compares the **spending decisions** of the most and least efficient schools and trusts, we find clear patterns of lower proportional spending across our two variables for the most efficient DMUs. The majority of the most efficient 100 schools and trusts spend less than ten per cent of their expenditure on 'back office' functions, compared with the majority of the least efficient DMUs spending over ten per cent, and about a half-a-dozen spending over twenty per cent of their expenditure on these functions. Similar patterns are seen for spending on education support staff. Nevertheless, we do see some of the most efficient schools and trusts with high proportional spends in these areas in line with the highest spends in the least efficient DMUs.

All in all, this analysis of the features of the most efficient DMUs has brought to light that these schools and trusts tend to have either substantially smaller leadership teams in relationship to pupil numbers or particularly low proportional expenditure on 'back office' functions, and to a slightly lesser extent lower spending on education support staff. The most efficient schools and trusts do not appear to employ teaching staff with significantly lower experience.

Figure 6.3: Full DEA results: Comparison of inputs and outputs between the top and bottom ranking DMUs. Staffing variables.

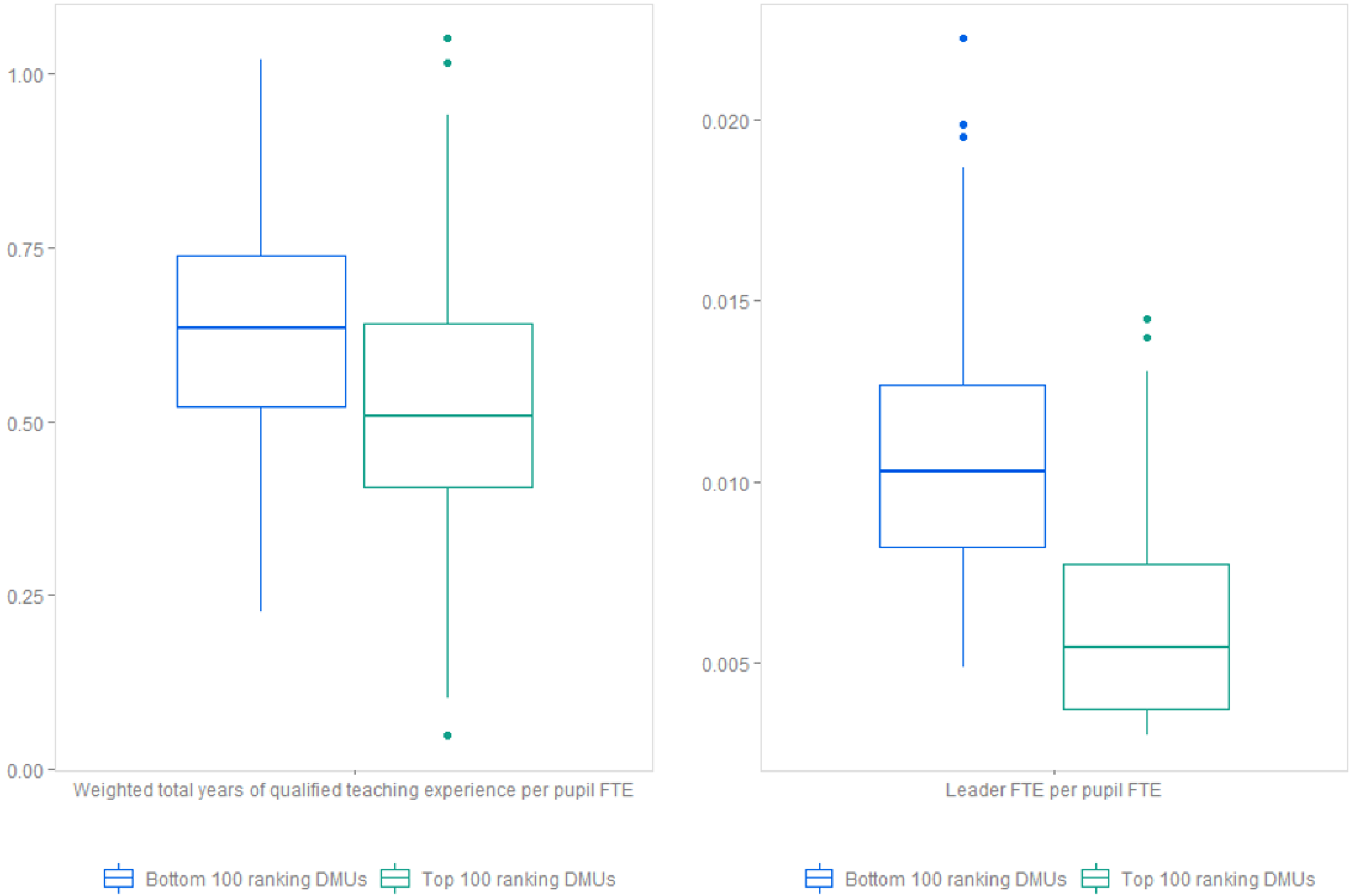
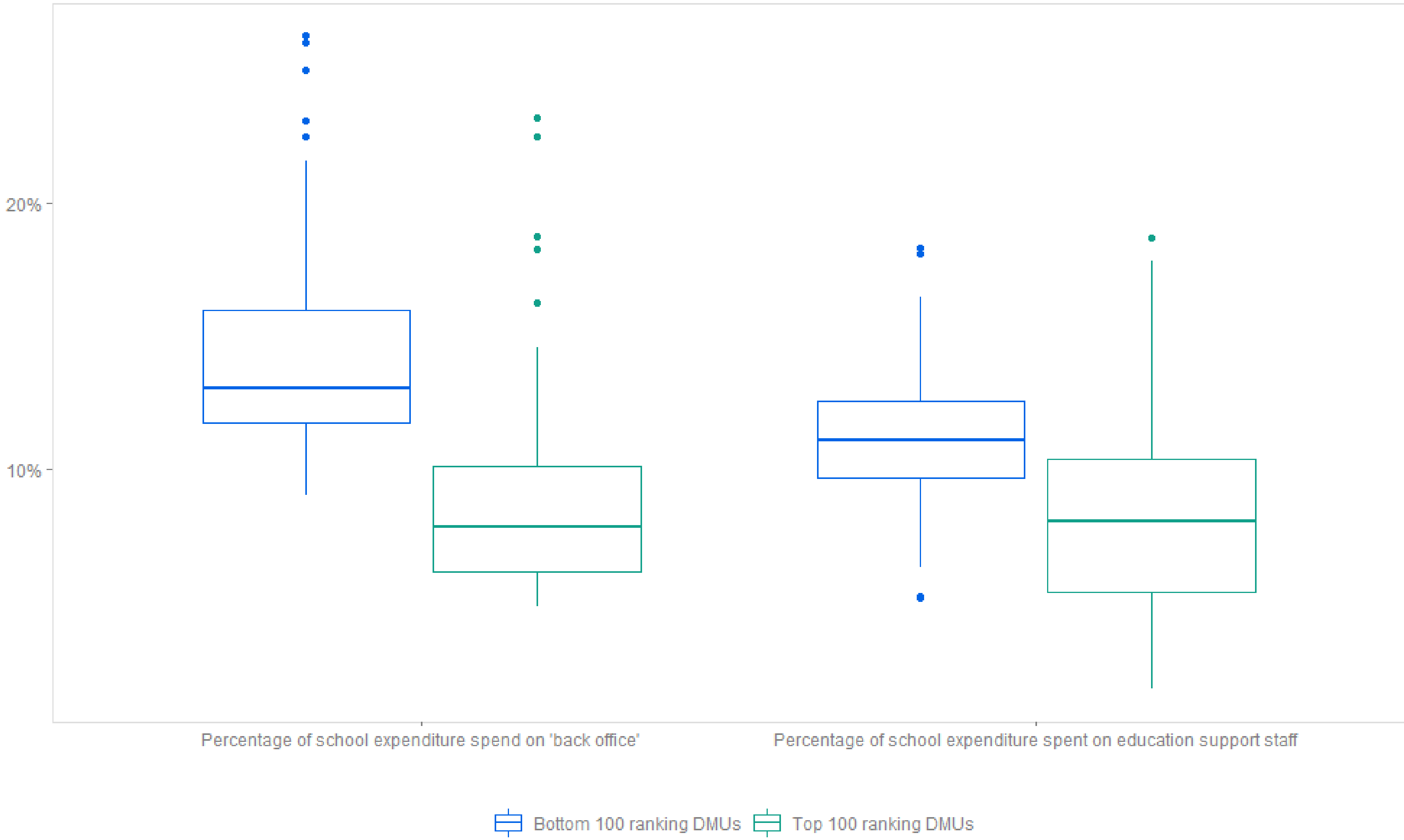


Figure 6.4: Full DEA results: Comparison of inputs and outputs between the top and bottom ranking DMUs. Spending variables.



7. Conclusion

This methodology paper has presented the foundational stages in developing our efficiency measures of school groups. We welcome and encourage feedback via the consultation.

Initial findings

Using our full four-input DEA model, we find that different types of school groups have broadly similar levels of efficiency to one another. Under this model, schools that are part of federations and foundation trusts appear to have comparatively low efficiency, whilst system leader trusts have the highest average efficiencies as a group. These differences between groups nevertheless cover a small range and evidence from this paper does not suggest that a particular type of school group is a more efficient operating model. Single-academy trusts, local authority schools and Diocese schools are most likely to be assessed as fully efficient by our model. Our model finds very few MATs to be fully efficient, however, we believe that multi-academy trusts are artificially less likely to be fully efficient under our model because, for these groups, any particularly high or low input values are 'averaged out' across multiple schools. An alternative would be to treat all schools as individual DMUs, however this might be a less faithful representation of some MATs. We invite views on this point through the consultation.

In analysing the features of the most and least efficient schools and trusts, we find the most significant differences to be the size of leadership team relative to pupil numbers, and lower spends on 'back office' functions as a proportion of total expenditure. Expenditure on education support staff is comparatively lower among the most efficient DMUs, however some of the most efficient schools still have similar spends in this area compared with the least efficient schools. In comparison with other input variables, there is striking overlap between the most and least efficient DMUs in terms of the level of teaching experience employed. Whilst teachers in the most efficient DMUs tend to have slightly fewer years of qualified teaching experience overall, it appears that high efficiency does not tend to be achieved with significantly less experienced teaching staff and greater efficiencies are found elsewhere.

These findings should not be interpreted as causal. For example, our finding that the most efficient schools tend to have smaller leadership teams relative to pupil numbers does not indicate that all schools should reduce the size of their leadership team in order to be more efficient. Rather, these findings prompt discussion and further research into when and how leadership teams should grow or be reduced, or whether instead of investing in more layers of leadership, funding might better be invested in developing the capacity of existing leaders. Our future qualitative research and surveys will take these questions forward.

Next steps for improving our measures

A key next step in improving this analysis is to add elements that will explore whether school efficiency is also both positive and sustainable. For example, do these efficient operating models deliver across a range of positive outcomes for pupils, such as pupil wellbeing, destinations, or achieving an appropriately broad curriculum? Further, are these staffing and spending practices sustainable for a workforce both at a group- and system-level, or do we detect concerning levels of staff turnover among the most efficient schools? Our future models must incorporate what are known as *negative outputs* in the DEA terminology, and also explore using multiple outputs and models that enable the assumption of *free disposability* to be relaxed. We could also look at other useful ways to segment our data, for example examining different school groups with similar numbers of pupils and comparing results that exclude schools in London.

In addition, future models will also benefit from improved input variables which better capture school operating models and resourcing. These inputs must better capture the key differences between DMUs which have an impact on pupil outcomes, and how resources are deployed between and within individual schools. For example, future models might include a broader range of spending and finance variables, such as spending on education resources, or long-term revenue reserve. Our leadership variable could be improved by capturing a dimension of quality or stability, rather than size of the leadership team. In the coming months we will be conducting qualitative and survey research which will add to our understanding of which elements are most important to capture in our input variables. Equally we invite views on our input variables via the consultation.

We must also ensure that our model fairly captures the contexts in which a DMU is operating, specifically in terms of the challenge of pupil intake and the historical performance profile of the school or trust. Our mechanism for accounting for context is the multi-level model which produces a predicted attainment score for each DMU based on the characteristics of its pupil intake, such as disadvantage, prior attainment, special educational need, first language, and ethnicity (see Figure 4.2 for full list of controls, and Appendix for model results). We will continue to improve on this model, for example by including historical school performance as a control for how well we would expect pupils to achieve given the improvement status of the school.

Finally, to improve our efficiency measures we must explore other methods for measuring efficiency, including Stochastic Frontier Analysis (SFA). During our analysis for this paper, we have attempted an SFA true random effects model as developed by Greene, assuming a truncated normal distribution for the inefficiency error term. [18,19] However under this model we found that the algorithm did not converge and therefore did not produce viable results. In future papers we will continue to build our understanding of this method.

[18] William Greene, “Reconsidering Heterogeneity in Panel Data Estimators of the Stochastic Frontier Model,” *Journal of Econometrics* 126, no. 2 (June 2005): 269–303, https://doi.org/10.1016/j.jeconom.2004.05.003; Tanja Kirjavainen, “Efficiency of Finnish General Upper Secondary Schools: An Application of Stochastic Frontier Analysis with Panel Data,” *Education Economics* 20, no. 4 (September 2012): 343–64, https://doi.org/10.1080/09645292.2010.510862.

[19] We use the same restricted panel as used in the full DEA model (i.e. outliers have been removed), limited to those DMUs where we have full data across all four years.

How to give feedback

We encourage and welcome your feedback in order to improve these measures of efficiency of school groups. Please return your feedback to this inbox feedback@epi.org.uk. The closing date for emailing feedback is 16th November 2021.

When you contact us, please provide us some details of who/which organisation you are representing with your views, for example a university faculty, a school or an academy trust.

Consultation questions

For school leaders and sector experts

- What suggestions would you make for improving our input variables? Can we improve our definitions of our existing variables? Are there any important inputs that are currently missing?
- In the paper we mention that we have selected a single output measure because the DEA tends to trade multiple outputs off one another as opposed to rewarding schools and trusts that achieve a balance across multiple outputs (page 14). Would you suggest any outputs that you believe can be acceptably traded-off against one another in education?
- Do you agree with our decision to treat MATs as a group and to take their inputs and outputs as an aggregate across all of their schools? What alternatives would you suggest?
- Do you see benefits of this analysis in addition to other available measures of efficiency such as the DfE's efficiency benchmarking tool?

For methodologists

- What advice do DEA experts have for relaxing the assumption of free disposability in this context?
- Do you have any comments on the methods applied here?

Our DEA model at a glance: Group efficiency of KS4 schools in England, 2018/19

We select four input variables (averaged across each year that the cohort was in the school, i.e. from Years 7 to 11)
Years of qualified teaching experience, totalled across all classroom teachers, per pupil
Number of leaders, per pupil
Proportion of total expenditure that is spent on 'back office' functions
Proportion of total expenditure that is spent on education support staff
We select one output variable
School 'value added'. This is calculated as a ratio of Attainment 8 scores to 'predicted' Attainment 8 for 2018/19. It represents the extent to which the school or trust enables its pupils to achieve higher than they would be expected to, on average, given their characteristics such as prior attainment and disadvantage.

Inputs and outputs are calculated for each *decision-making unit* (DMU).
Where schools are part of an **academy group**, inputs and outputs are calculated at **group-level. The decision-making unit is therefore the academy trust for all types of multi-academy trusts (MAT).**
Where schools are **standalone** or part of a **non-academised group** (such as a federation, diocese or foundation), inputs and outputs are calculated at **school-level. The decision-making unit is therefore the school for these types of group.**
For all groups, we only include schools that have been part of their current group for at least two years by the beginning of the 2018/19 academic year. DMUs with improbably high or low values are removed from the data.

The resulting dataset is used to in the DEA calculation
The DEA calculation is output-oriented and assumes constant returns to scale.
The calculation produces efficiency scores for each DMU. To find median efficiencies by group type, we first find the geometric mean efficiency score of individual groups with multiple DMUs. For example, for a specific Local Authority, we find the geometric mean of the efficiency scores of all schools (DMUs) in that Local Authority. This gives a single efficiency score for each individual group. We then take the median of these by group type.

Appendix

A: Results of multi-level model predicting pupil Attainment 8 account for various pupil and school characteristics

effect	group	term	estimate	standard error	statistic	95 per cent confidence interval - lower	95 per cent confidence interval - higher
fixed		Intercept	73.77	2.51	29.34	68.84	78.70
fixed		Prior attainment - KS2 average point score value-added variable	-10.12	0.24	-42.68	-10.58	-9.66
fixed		Prior attainment squared	0.28	0.01	28.67	0.26	0.30
fixed		Prior attainment cubed	0.00	0.00	-7.32	0.00	0.00
fixed		Pupil is male	-4.44	0.04	-123.49	-4.51	-4.37
fixed		Pupil has any identified special educational need or disability (SEND)	-6.79	0.06	-121.15	-6.90	-6.68
fixed		Pupil is disadvantaged (EVER6FSM)	-4.54	0.04	-104.17	-4.62	-4.45
fixed		IDACI Score of pupil's postcode	-13.24	0.17	-76.88	-13.58	-12.91
fixed		IDACI Score is missing	3.85	0.99	3.90	1.92	5.79
fixed		Pupil not in state system at KS1	5.82	0.08	72.53	5.67	5.98
fixed		Pupil speaks English as an additional language (EAL) (ref = English as first language)	0.04	0.11	0.34	-0.18	0.26
fixed		Pupil's language group is unclassified (ref = English as first language)	0.44	0.75	0.59	-1.02	1.90
fixed		Major ethnicity code - Any other ethnic group	2.61	0.14	18.78	2.34	2.89
fixed		Major ethnicity code - Asian	4.50	0.08	59.93	4.35	4.65
fixed		Major ethnicity code - Black	2.36	0.08	27.78	2.19	2.53
fixed		Major ethnicity code - Chinese	7.99	0.28	28.40	7.44	8.54
fixed		Major ethnicity code - Mixed	1.56	0.08	19.39	1.40	1.72
fixed		Major ethnicity code - Unclassified	-0.29	0.16	-1.80	-0.61	0.03
fixed		Pupil lives in London	1.20	0.15	8.20	0.92	1.49

fixed		School level - average prior attainment	1.71	0.05	33.78	1.61	1.81
fixed		School level - standard deviation of prior attainment	2.32	0.10	23.48	2.13	2.51
fixed		Interaction - IDACI Score and missing IDACI flag	-20.68	4.56	-4.53	-29.62	-11.74
fixed		Interaction - IDACI Score and pupil is EAL	6.55	0.37	17.93	5.83	7.27
fixed		Interaction - IDACI Score and pupil's first language is unclassified	-0.22	2.68	-0.08	-5.47	5.02
ran_pars	ks4_urn	NA	3.63	NA	NA	NA	NA
ran_pars	Residual	NA	12.27	NA	NA	NA	NA

Source: National Pupil Database 2018/19 KS4, School Census 2019, Department for Education

B: Correlation of inputs and outputs

	Average Attainment 8	Value added - Attainment 8 divided by predicted attainment	Predicted Attainment 8, contextualised	FTE of Leaders, per pupil FTE	Total years of qualified teaching experience in the DMU, weighted by FTE of experienced qualified teachers, per pupil FTE	% of total expenditure spent on education support staff	% of total expenditure NOT spent on back office functions
Average Attainment 8	1.0	0.4	0.9	-0.3	0.2	-0.4	0.0
Value added - Attainment 8 divided by predicted attainment	0.4	1.0	0.0	0.1	-0.1	-0.1	0.0
Predicted Attainment 8, contextualised	0.9	0.0	1.0	-0.4	0.3	-0.4	0.1
FTE of Leaders, per pupil FTE	-0.3	0.1	-0.4	1.0	-0.3	0.1	-0.2
Total years of qualified teaching experience in the DMU, weighted by FTE of experienced qualified teachers, per pupil FTE	0.2	-0.1	0.3	-0.3	1.0	-0.1	0.2

% of total expenditure spent on education support staff	-0.4	-0.1	-0.4	0.1	-0.1	1.0	0.2
% of total expenditure NOT spent on back office functions	0.0	0.0	0.1	-0.2	0.2	0.2	1.0