

Teaching Improvement through Data and Evaluation (TIDE)

Annual report 2026

Understanding, measuring and improving teaching impact

NATIONAL INSTITUTE of Teaching

Table of contents

Foreword	4
Executive summary	5
The team	8
Introduction	10
1. Can we estimate teaching impact?	11
2. What factors are associated with teaching impact?	21
3. Can we improve teaching impact?	26
4. Responsible use of data and AI	29
5. Case Study – Oasis Community Learning	32
6. What next for TIDE?	35
Get involved	36
Bibliography	37
Appendices	40
Our Strategic Advisory Board (SAB)	41
Our Expert Working Group (EWG)	42

Nuffield Foundation

The Nuffield Foundation is an independent charitable trust with a mission to advance social well-being. It funds and undertakes rigorous research, encourages innovation and supports the use of sound evidence to inform social and economic policy, and improve people's lives. The Nuffield Foundation is the founder and co-funder of the Nuffield Council on Bioethics, the Ada Lovelace Institute and the Nuffield Family Justice Observatory.

This project has been funded by the Nuffield Foundation, but the views expressed are those of the authors and not necessarily the Foundation.

Find out more at: nuffieldfoundation.org

Bluesky: @nuffieldfoundation.org

LinkedIn: Nuffield Foundation

Foreword

Everyone who works in schools knows how important teachers are. An effective teacher can change the course of a child's life. Experienced school leaders usually have a good sense of who their most impactful teachers are, but that knowledge is tacit and decentralised. We do not have a reliable way of measuring it. That becomes a problem when we try to understand where teaching expertise comes from, and how to grow more of it. We cannot easily say how much teachers vary in their impact, what the most effective teachers do differently, or whether the large sums we spend developing them actually work.

That gap has always struck me as odd. In almost every other field, we measure what matters and use what we learn to improve. In teaching, the thing that matters most has gone largely unmeasured. Not because it cannot be done, but because the data have been too fragmented, and the tools too blunt, to do it well.

That is why I am proud that the Harris Federation is one of the founding trusts behind TIDE. Like the other founding trusts, we run common assessments across many schools, which means for the first time we can compare and learn at a scale that simply was not possible before. The National Institute of Teaching has taken that opportunity and built something genuinely new. A dataset, and a way of working with it, that lets us ask serious questions about how teaching develops and how pupils learn.

I want to be clear about what this is, and what it is not. This work is not about judging or ranking individual teachers. The data are anonymised, and they are designed to help us understand teaching in general, so that we can train and support our teachers better. It would be a mistake, and a betrayal of the profession's trust, to use it any other way. The report you are about to read is careful on exactly this point, and rightly so.

Since the project began in July 2024, it has shown the real potential of this approach. It seems that we can estimate teaching impact, and begin to see how it is distributed within a trust, within a school, and even across the different classes a single teacher takes. We are starting to understand what lies behind it. There is a great deal still to do, and the team is honest about that. But the foundations are now in place, and the potential is considerable. If this work succeeds, it will help not only our own schools but the whole sector to develop teachers more effectively, and to give every child the teaching they deserve.

I am grateful to the Nuffield Foundation for making this possible, and to the teachers and leaders across our trusts whose work this builds on. I look forward to seeing where it leads.

Sir Dan Moynihan

Sir Dan Moynihan
Chief Executive, Harris Federation

Executive summary

Research that links pupils to their teachers consistently finds that teachers differ markedly in their impact on attainment, and that this is among the most important in-school influences on pupil outcomes (Chetty et al., 2014; Hanushek and Rivkin, 2010). However, most of that evidence has come from abroad, perhaps because England has lacked a reliable way to measure teaching impact at scale. This has made it challenging to build an evidence base on teacher effectiveness and its drivers, a gap that the Teaching Improvement through Data and Evaluation (TIDE) project seeks to address.

1. Can we estimate teaching impact?

Working with the National Institute of Teaching's (NIoT) founding multi-academy trusts (MATs), we find that anonymised pupil assessment data can be used to estimate teacher impact, which we measure using an approach known as teacher value-added (TVA).

There are, of course, many contributors to a pupil's assessment result aside from their teacher that year. Teacher value-added is designed to isolate the teacher's own distinct contribution to a pupil's progress, separately from these other influences. Our models therefore capture as many of those other factors as possible, such as their prior attainment, current attainment in other subjects, demographic characteristics (including their sex, eligibility for free school meals, special educational needs, and whether English is their first language). However, our data cannot capture everything, and pupil attainment on assessments will also be subject to some random variation, just like for any human being completing any task. Furthermore, our estimates of teaching impact assume the assessments are of sufficient quality, that assessment content and conditions are standardised across the MAT, and the questions must target the curriculum precisely.

We are satisfied that in our analysis these conditions have, for the most part, been met. Our participating MATs administer standardised and high-quality assessments across entire cohorts (sometimes thousands of children) in similar conditions, with all children following the same, or very similar, curriculum. By analysing actual pupil performance compared to predictions based on prior attainment and their demographics, our models can estimate the impact of their teacher. We have shown that this is feasible not just in primary schools (where each teacher typically has just one class) but also secondary (where a teacher usually has multiple classes). We also found that teachers and leaders find analysing assessment data in this way acceptable, as long as the purpose is to conduct research that might contribute to system-level professional development, and not performance management of individual teachers.

Our estimates of teacher impact are not perfect, as we have incomplete data and some assessments have room for improvement. Moreover, our estimates would be imperfect even with complete data and ideal assessments, as pupil performance will vary for reasons beyond a teacher's control. We therefore cannot usefully compare one individual teacher directly with another. However, our analysis does let us describe, for the first time in England, how teaching impact is distributed within and between schools, which may shed light on the drivers of educational inequality and how to reduce it. This work is published on the NIoT's **website**, alongside a fuller technical **working paper**.

2. What factors are associated with teaching impact?

We have begun exploring whether our measure of teacher impact is associated with a teacher's training pathway, allocation of classes, experience, observed classroom practices, or the professional culture of their school. We are still assembling the data necessary to do this, but have made good progress using schools' timetable data, HR records, surveys, and records of peer observations. Three early findings stand out:

- First, an analysis of within-year class moves suggests that moving pupils between sets mid-year tends to carry a small net cost to attainment, implying the disruption of changing classes mid-year might be more harmful for attainment than any potential efficiency gain from being taught with more similar pupils.
- Second, the Department for Education (DfE) asked us to explore whether pupils with special educational needs and disabilities (SEND) were allocated less impactful teachers. We tested this hypothesis in one MAT and found no evidence this was the case. We will replicate this study with more MATs in future.
- Third, we have found that it may be possible to identify specific teaching practices from an anonymised transcript of a lesson. In our pilot, several Early Career Teachers (ECTs) and their mentors recorded approximately 30 lessons between them using a secure recording tool. NIoT tutors then assessed whether 29 distinct teaching practices were present or not. We then analysed the extent of agreement between tutors to assess which practices are likely to be most feasible for labelling in this way at scale.

Given that our models and data cleaning processes are still at a developmental stage, our findings here should be treated with caution, as they are associations rather than causal relationships. However, we hope they will contribute to MAT or teacher development provider-level innovations that can be evaluated more robustly in future. TIDE will develop and evaluate the impact of two such innovations, which are described in the next section.

3. Can we improve teaching impact?

In 2027–28 we will use our insights to develop and evaluate interventions designed to improve teaching impact. We will evaluate the impact of AI assistance for mentors observing ECTs in a randomised controlled trial (RCT) and publish our trial protocol in advance. We are also exploring whether our recording tool could be used for peer observation. Lastly, we will explore whether any other insights arising from the work described above (such as changes to timetabling or ability grouping) could inform innovations that can be evaluated causally.

Meanwhile, we have developed and shared prototype bespoke 'MAT reports' presenting each trust with its own analysis of teaching impact. We are in the process of gathering their feedback so we can make these reports more useful in future. Again, these insights are anonymised at the teacher level, ensuring that they support general developmental practices rather than individual performance management.

4. Responsible use of data and AI

TIDE works with sensitive data, and handling it responsibly is a condition of the project's legitimacy. The report sets out how the Teacher Education Dataset (TED) brings school data together in pseudonymised form, and how the TED will be accessed by researchers in future via OpenSAFELY-Schools, ensuring secure and reproducible analysis.
We also set out the extensive legal and ethical groundwork underpinning the recording and analysis of lessons in the Analysis of Teaching Practice (ATP) project.
Throughout, teacher-impact estimates are developmental tools, never instruments of performance management.

5. What next for TIDE?

Over the 2026–27 school year we will refine our estimates as we receive additional annual tranches of data (i.e. attainment from 2025–26) and integrate new data sources. This will allow us to investigate the factors associated with teaching impact and report findings by the end of 2027. We will also develop the trial protocols.

In 2027–28, the (ATP) project will move into its evaluation phase, with an RCT of AI-supported mentoring for ECTs. By the end of 2028, we aim to have completed the research, published its outputs, and contributed an infrastructure and a way of working that allow us and others to keep improving teacher development at scale, long after TIDE has concluded.

Get involved

TIDE depends on the trusts that contribute their data, and we are keen to work with more. If you would be interested in contributing data to the TED, or in using it for your own research, we would be glad to hear from you.
Data: **ted@niot.org.uk**
Research: **research@niot.org.uk**

The team

Dr Calum Davey
Executive Director of Research and Best Practice at the NIoT, and Principal Investigator on TIDE. Calum has extensive expertise in the design and analysis of large-scale randomised controlled trials, developed across public health and international development and, before joining the NIoT, at the London School of Hygiene and Tropical Medicine. He sits on the Education Endowment Foundation's (EEF) Evaluation Advisory Group, and manages the NIoT's £8millon DfE research portfolio.

Professor Rob Coe
Director of Research and Development at Evidence Based Education and a Senior Associate at the EEF. Rob is one of the country's leading authorities on what makes teaching effective and how research reaches the classroom, reflected in the Great Teaching Toolkit and his widely cited review 'What makes great teaching?'. Rob has led the development of the TVA measure which underpins this project.

Professor John Jerrim
Professor of Education and Social Statistics at University College London (UCL), and Director of its Quantitative Social Science Research Centre. John is an expert in large-scale education datasets, having led England's work on PISA and TALIS and a major Nuffield-funded study of teacher mental health and wellbeing. He publishes extensively on socio-economic inequality in education. John leads our work on the associations between class characteristics and TVA.

Dr Sam Sims
Associate Professor at UCL and Director of Research at StepLab, whose expertise spans teacher recruitment and retention, and the methodology of causal inference in education. Sam led the EEF's systematic review of the characteristics of effective teacher professional development and works widely with large administrative datasets. Sam leads our work on the association between a school subject department's social capital and the TVA of its teachers.

Professor Becky Allen
Co-founder of TeacherTapp and Professor of Education at the University of Brighton, Becky leads our work on the association between training pathways and TVA.

Dr Raj Chande
Senior Research Associate at the NIoT, and until recently a maths and economics teacher at a secondary school in Hackney. Raj combines classroom practice with a research career leading large-scale education RCTs for the EEF and DfE. Raj supports Calum across all aspects of the TIDE project.

Claire Maud
Research Fellow at the NIoT. Claire pairs 18 years of experience in secondary schools with a recently completed master's in computer science, bringing together classroom knowledge and technical data skills. Claire leads on the MAT reports product and our work on the association between TVA and observed teaching practices.

Dr Bronte McKeown
Senior Research Fellow at the NIoT, Bronte leads the ATP project. Bronte's background is in psychology and cognitive neuroscience, with expertise in machine learning and data science.

Dr Rachel Conlon-Bose
Research Statistician at Evidence Based Education, Rachel supports Rob Coe.

Shaun Dillon
Shaun is the NIoT's Director of Data and Data Protection Officer. Shaun leads on obtaining and cleaning data from participating MATs.

TIDE is supported by a Strategic Advisory Board (SAB), which provides independent advice, support and critical challenge, and an Expert Working Group (EWG) of experienced teachers and teacher-development experts, who ground the research in the realities of schools. Membership of both is listed in the appendices.

Introduction

Teaching quality is widely recognised as one of the most important in-school factors influencing pupils' educational outcomes. Every year, substantial resources are invested in recruiting, training and developing teachers, yet we still know surprisingly little about which approaches are most effective at improving teaching and, ultimately, pupil attainment. One of the central barriers has been the absence of the data and research infrastructure needed to study teaching quality at scale.

The Teaching Improvement through Data and Evaluation (TIDE) project was established to address this challenge. Funded by the Nuffield Foundation in July 2024 and led by the NIoT in partnership with leading researchers and sector organisations, TIDE seeks to improve outcomes for pupils by developing our understanding of teaching quality and how it can be developed. At the heart of the programme is the Teacher Education Dataset (TED), a unique collaboration bringing together pupil, teacher and school data from participating MATs, alongside new forms of evidence about classroom practice and professional development. The project is organised around three broad questions:

1. Can we estimate teaching impact?

2. What factors are associated with teaching impact?

3. Can we improve teaching impact?

These questions reflect a simple goal: to build a stronger evidence base for improving teaching and learning across the education system. This is the first in a series of annual reports that will document TIDE's progress and emerging findings as the programme develops. The project is ongoing, with analysis, development and evaluation work still to come, with the final report published by the end of 2028.

The foundation for this work is the TED, which connects pupils' assessment results to the teachers who taught them, alongside pupil prior attainment and background characteristics and information about teachers themselves. It currently draws on data from 7 MATs, spanning 91 primary, 124 secondary and 7 all-through schools across every region of England. In total it covers around 165,000 pupils, 8,500 teachers and 2.5 million assessment results. This is already beyond the scale we set out to reach at the start of the project, and it continues to grow as further trusts join and each new year of data is added. The data are pseudonymised so that no individual school, teacher or pupil is identifiable in our analysis.

1.
Can we estimate teaching impact?

The 'value-added' impact of schools has long been a focus of researchers and policymakers (Koedel et al., 2015). More recently, researchers have also explored the impact of individual teachers, but most of that evidence has come from abroad, perhaps because England has lacked a reliable way to measure teacher impact at scale. TIDE seeks to address that gap.

1.1 What do we want to know?

Research linking pupils to their teachers shows that teachers vary widely in their impact on attainment, and that this variation matters more than almost any other in-school factor (Chetty et al., 2014; Hanushek and Rivkin, 2010). Every year, schools in England spend nearly £3,000 per teacher on professional development, once staff time is included (EPI, 2021). However, few datasets link pupil assessment results to their teachers, so while teacher development programmes often ask participants if they found their courses worthwhile, we have little reliable evidence as to which approaches to training and development have the greatest positive effect on pupil attainment. Building an evidence base for teacher training and continuing professional development (CPD) depends, first, on being able to measure teaching impact.

Throughout this report, when we refer to a teacher's impact we mean their estimated contribution to their pupils' progress, measured using an approach known as teacher value-added (TVA). Our central question in this strand of TIDE is whether it is possible to estimate teachers' impact on pupil attainment using data that schools already collect. Researchers in the United States have done so by comparing the progress of a teacher's pupils against similar pupils taught by others (e.g. Kane et al., 2011; Chetty et al., 2014), an approach reviewed extensively in the value-added literature (Koedel et al., 2015). This has not previously been feasible at scale in England because it requires pupils to sit the same assessment at the same time, and because these data are not linked to individual teachers. Aside from GCSEs and Key Stage 2 SATs this rarely happens, which makes comparison across teachers in different schools impractical (Coe et al., 2024). One notable exception used this approach to identify high value-added Year 11 teachers (Burgess et al., 2022). Assessments in England are, however, nationally standardised only in Years 6 and 11, so teachers of other year groups cannot readily be compared across schools. The growth of large MATs changes this. Within a single trust, thousands of pupils across dozens of schools may sit common, internally standardised assessments, creating for the first time the conditions under which TVA can be estimated across phases and year groups in England (Coe et al., 2024).

We wanted to establish not only whether such estimation is possible, but also whether using anonymised pupil assessment data linked to teachers for research purposes is acceptable to teachers and school leaders. In addition, we wanted to explore whether the data could be obtained in a usable form, whether school assessments are of sufficient quality, and whether our models recover genuine differences in teacher impact rather than just random variation in pupil attainment (sometimes called 'statistical noise') that has nothing to do with the teacher (Coe et al., 2024).

Two methodological concerns are well recognised in this literature. The first is bias. If classes are allocated to teachers based on characteristics not captured in our data, then any measure of teacher effectiveness based on pupil performance might be inaccurate. For example, it might be that a head of department allocates themselves more challenging (or easier) classes in terms of behaviour. If that class then underperforms (or outperforms) expectations based on prior attainment, then it would be inaccurate to attribute all of that difference to the effectiveness of the teacher. Reassuringly, studies that exploit random assignment or track teachers across schools find value-added estimates are largely stable (Kane and Staiger, 2008; Chetty et al., 2014). The second is noise, since estimates based on small numbers of pupils can be imprecise. This can be mitigated by shrinking estimates towards the mean in proportion to the amount of data behind them (Koedel et al. 2015).

Crucially, this work is about understanding the development pathways and practices associated with high-impact teaching, so that teacher education and continuing professional development (CPD) can be made more effective. It is not an exercise in ranking or performance-managing individual teachers. The estimates are strictly anonymised, with no individual teacher, school or pupil identifiable. Measuring teaching impact is a means to an end. The purpose is to improve how teachers develop, and ultimately to improve outcomes for pupils.

Working with our founding MATs, we therefore set out to answer four questions:

1. **Is it acceptable to teachers and school leaders to use anonymised assessment data for this purpose?**
2. **Is it feasible to obtain the data in the format we need?**
3. **Are the assessment data of sufficient quality to support this kind of research?**
4. **Does our analytical approach show that teacher impact can be estimated, or do the results look like statistical noise?**

1.2 What have we learned so far?

Working with our founding MATs, we combined qualitative interviews with trust leaders and teachers with quantitative analysis of their anonymised assessment data. Our findings against the four questions are encouraging.

Is it acceptable to teachers and school leaders?

We found broad acceptance among teachers and leaders for using anonymised assessment data in this way, provided its purpose is to identify effective development pathways rather than to judge individuals.

Since sharing this work, some MAT leaders have asked us how they might find out which of their teachers are the most effective, so that others could learn from them. We cannot do this, as doing so would violate the terms under which the data were shared with us. We do, however, understand the instinct behind the question. Wanting to know what the most effective teachers do differently is exactly what drives the next phase of our work. Research questions 2d and 2e, which examine lesson observations and recordings of teaching, are designed to answer that question at the level of practice rather than the individual, so that what effective teachers do can be shared without any teacher being identified or ranked.

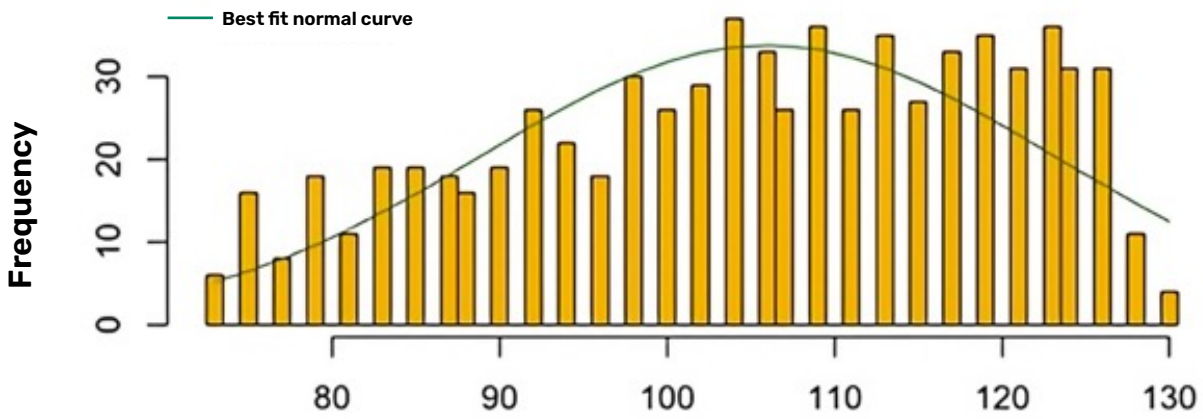
Is it feasible to obtain the data?

We found that collecting and matching the data is feasible, but it is complex and demanding. The core difficulty is that school management information systems are not designed to link pupil assessment results to the teacher who taught them. Making that link requires working through timetables to establish who taught which pupils in which class: a task that is both intricate and time-consuming. Assessments are sometimes recorded in inconsistent ways within a single trust, and information often has to be drawn together from several different datasets. The result is a substantial amount of linking and cleaning before any analysis can begin.

Are the assessment data of sufficient quality?

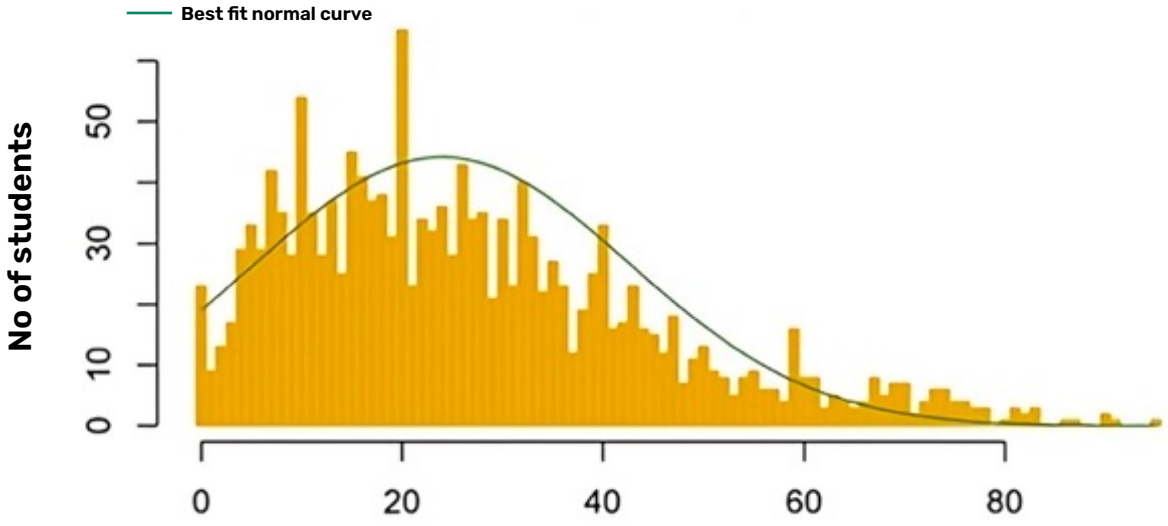
We found that most assessments are of sufficient quality for our intended analyses, though some are better suited than others. The three distributions in Figures 1 to 3 illustrate this. An assessment can be too easy, where too many pupils score at or near the top and their scores bunch up against a ceiling, so the test cannot distinguish between them (a 'ceiling effect', Figure 1).

Figure 1: Assessment that is too easy (ceiling effect)



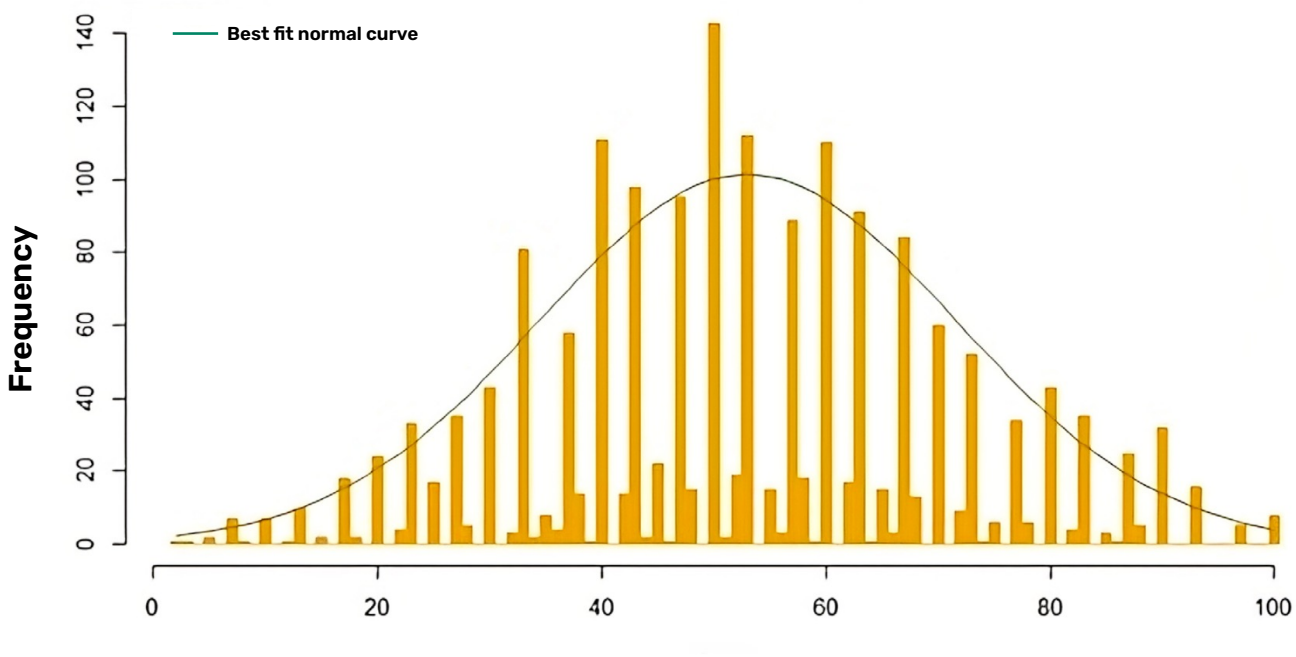
An assessment can be too hard, where too many pupils score at or near the bottom against a floor, with the same loss of information (a floor effect', Figure 2).

Figure 2: Assessment that is too hard (floor effect)



A well-designed assessment spreads scores across the range, as in Figure 3, which allows genuine differences in attainment to show through. While a minority of assessments could be improved, most were found to be suitable for estimating teaching impact.

Figure 3: Assessment with a good (normal) distribution of pupil scores



We also checked the assessments in a second way. For a given pupil in a given subject, we do not expect their performance to swing wildly from one point in the year to the next. A pupil's maths score in the autumn and in the summer should be related, because underlying attainment is reasonably stable over a year. Finding that kind of expected stability within pupils, across assessments taken at different times, gives us confidence that the assessments are measuring something real rather than producing noise. Arguably the most important application of this measure of validity is the extent to which an assessment predicts final GCSE attainment. Figures 4 and 5 below show the relationship between a pupil's result in an assessment in Year 10 or 11 (x-axis) and their GCSE grade (y-axis) in a specific MAT. Figure 4 (top) shows the relationship for maths. The green lines have been added to illustrate the range of data around an approximate line of best fit. Figure 5 (bottom) shows the relationship for English, with the same green lines to illustrate the weaker correlation for English. This shows that English assessment results are less predictive of GCSE grades than is the case for maths. Indeed, we find there to be more stability from one assessment to the next in maths than for other subjects, likely because of its relatively objective marking criteria.

Figure 4: Year 10 maths assessment vs GCSE grade

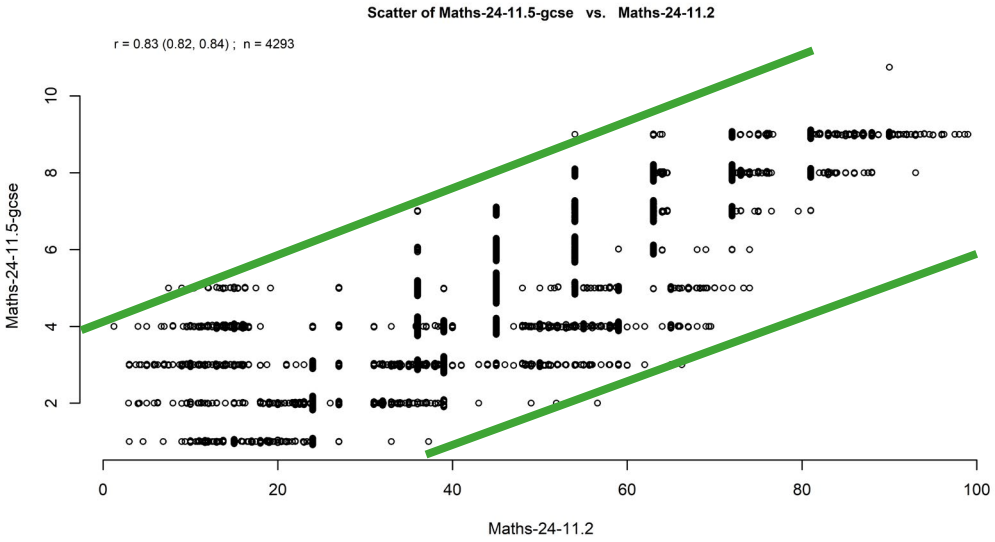
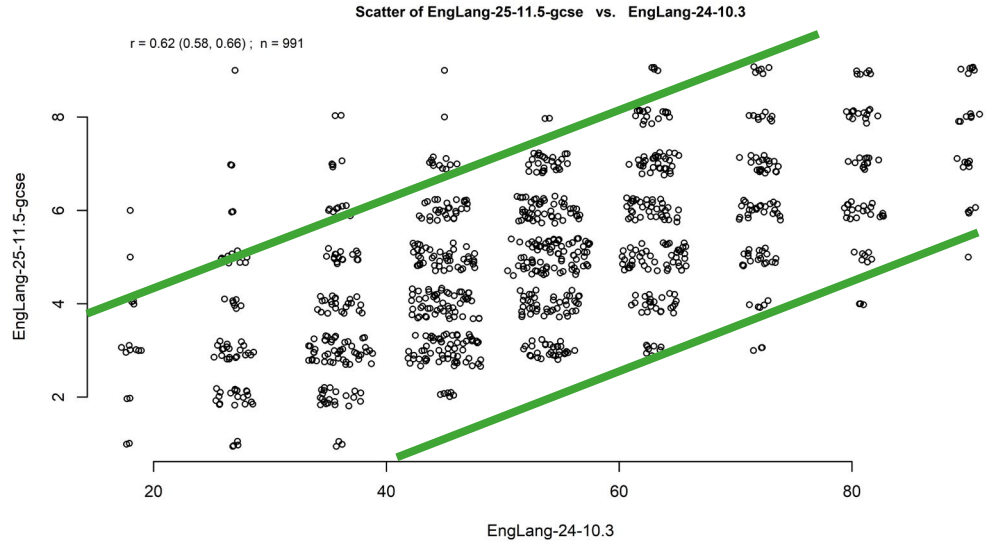


Figure 5: Year 10 English assessment vs GCSE grade



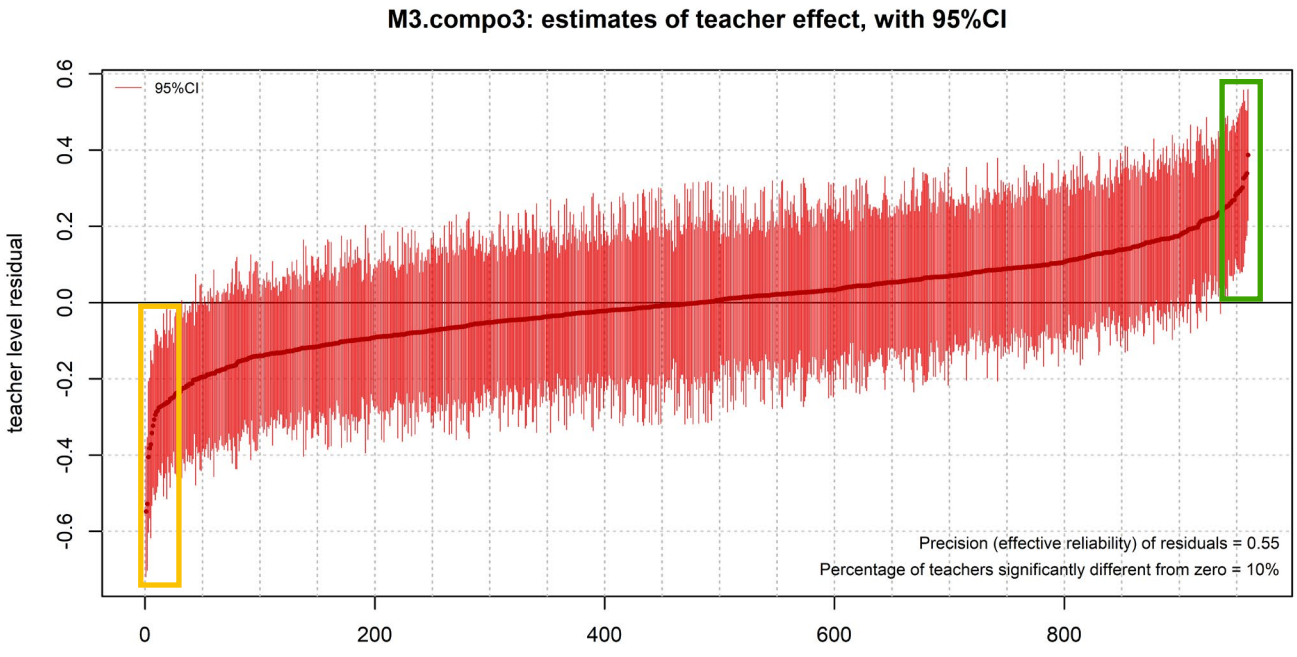
This quality check has also proved useful to the trusts themselves. While most assessments were of sufficient quality, the same analysis let us identify specific assessments that could be improved: some were poorly calibrated, with scores bunching at the top or bottom, some were inconsistent from one year to the next, and some related only weakly to later attainment. This matters beyond our research, because an assessment that does not track pupils' progress reliably is of limited use to a school. Feeding this back to trusts, so they can interrogate the quality and purpose of their own assessments, has been an early and tangible benefit of the work.

Can teacher impact be estimated, or is it just noise?

We found that our analysis produces estimates of teaching impact that are meaningful rather than random statistical variation. Three observations support this.

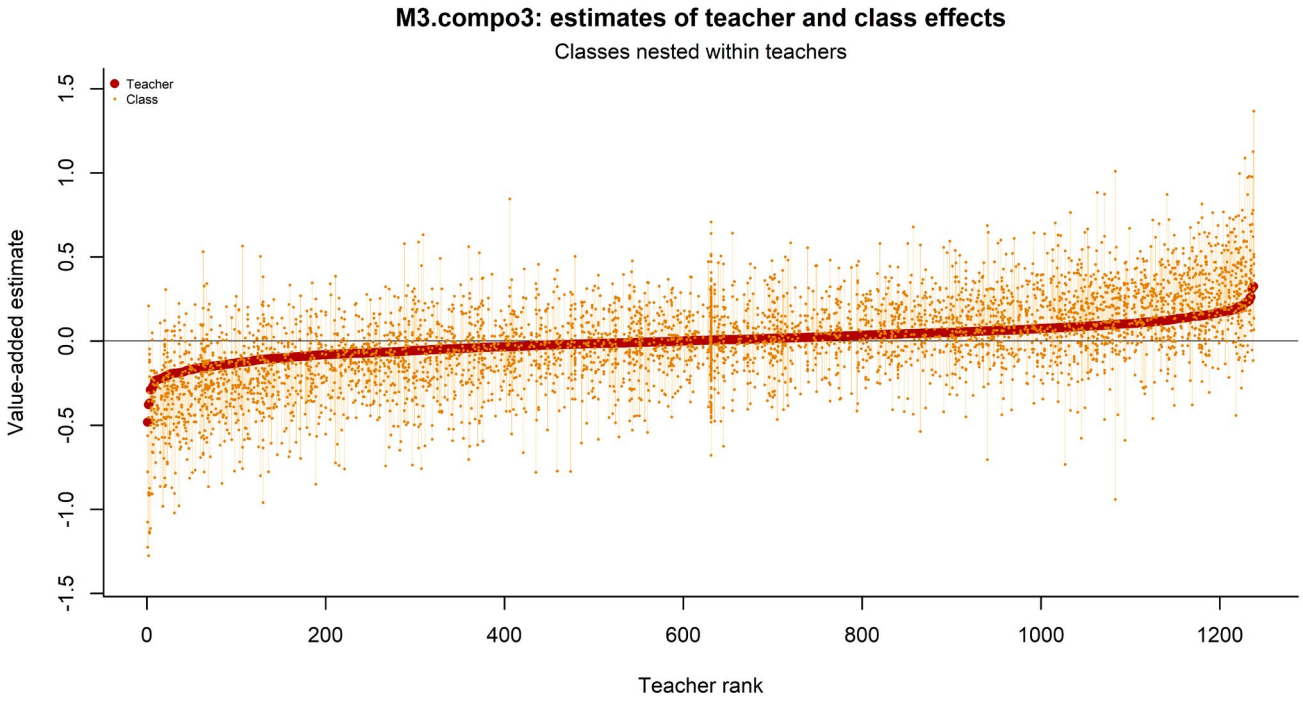
First, we can describe the distribution of teaching impact across a MAT. Because the analysis relies on pupils sitting common assessments, and those assessments are standardised within a trust rather than across trusts, we analyse one trust at a time. Figure 6 shows approximately 1,200 teachers in one MAT, ranked (horizontal axis) from lowest to highest value-added (vertical axis), using the model most commonly used by researchers. The confidence intervals, shown as the vertical lines around each estimate, cross zero for 96% of teachers, which means we cannot say with confidence that the pupils of any given teacher in this group did better or worse than expected. Over time, as we accumulate more data per teacher, the confidence intervals might narrow, allowing our model to identify variation in value-added with more confidence. For now, only a relatively small number of teachers (green box) outperform expectations by a statistically significant margin, while a similarly small number underperform (orange box). This finding underlines the importance of not using our measure to performance-manage any individual teacher, but instead to analyse the distribution of teacher effectiveness across an entire MAT.

Figure 6: Distribution of estimated teacher value-added in a single MAT



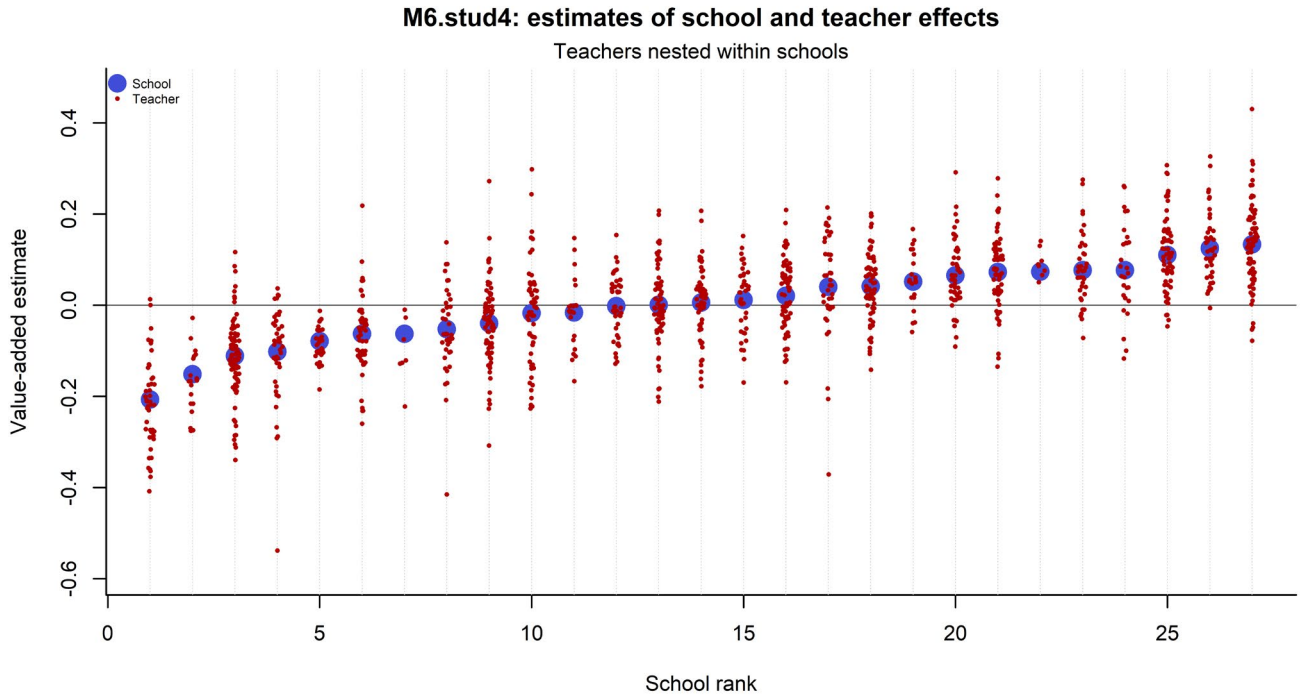
Second, a teacher's impact varies across the classes they teach. Figure 7 shows the range of estimates for each teacher across their different classes. Some teachers with lower value-added overall have nonetheless had a positive impact with some of their classes, and the reverse is also true. Teachers whose classes perform in uniformly positive or uniformly negative ways are rare. This within-teacher variation reinforces the point that a single number cannot capture the impact of an individual teacher, as so much depends on the particular class. What we can say with more confidence relates to the broader pattern: across all their classes, some teachers tend to outperform expectations and others tend to underperform them.

Figure 7: Estimates of class effects (orange) clustered by teacher (red)



Third, and more exploratory, teaching impact varies within schools as well as between them. Figure 8 shows that while some schools have a higher average impact than others, each school's average is made up of considerable variation between the individual teachers within it, and this varies more in some schools than others. We treat this as an early, exploratory observation and look forward to examining this variation further in future work.

Figure 8: Estimates of teacher value-added (red) clustered by school (blue)



The work described so far is drawn mainly from secondary schools, where the approach is most tractable. We have also estimated teaching impact in primary schools, though this rests on a much smaller sample of 79 teachers across 13 schools, and is more challenging (Coe et al., 2024; Jerrim et al., 2025; Allen et al., 2026). The central difficulty is that primary teachers typically teach one class across several subjects, rather than several classes in one subject, as in secondary. This means the teacher and the class are largely the same unit, which makes it harder to separate a teacher's contribution from the particular mix of pupils they teach. It also removes one of the checks available in secondary, where we can compare a teacher's value-added across their different classes. In primary we instead look at consistency across subjects, such as maths and reading, for the same teacher. With fewer teachers and schools in the primary sample, the estimates are also less precise. Investigation in primary schools therefore remains an area we are developing.

This work is now published in full on the NIoT's **website**, alongside Rob Coe and colleagues' more extensive working paper with its complete technical appendix.

1.3 What happens next?

The immediate priority is to keep refining our estimates of teaching impact as more and richer data become available. For example, we previously made a simplifying assumption that if multiple teachers are linked to the same class, then those teachers have an equal amount of time and therefore opportunity to impact the attainment of that class. We have now obtained timetable data from two of our founding MATs, which tells us more precisely how long each class was taught by a particular teacher, and we will use this to run improved value-added analyses. Furthermore, each additional year of attainment data will allow us to sharpen the estimates and narrow the confidence intervals, moving us closer to being able to distinguish between higher- and lower-impact teaching.

A key shift over the next phase concerns how we obtain that data. Our analysis so far has relied on data leads within MATs manually extracting and pseudonymising data on their side. This work is demanding for trust staff, and difficult to scale. As the TED infrastructure is built, we will progressively move from these manual extracts towards drawing data directly from MATs' management information systems (MIS; e.g. Bromcom) via an Application Programming Interface (API). This will reduce the burden on trusts, make the process more consistent, reproducible, and secure, and allow the research to extend more easily to further MATs over time.

We are also currently co-creating bespoke insights for MATs contributing their data. We call this product 'MAT reports'. We are providing the analyses above for individual MATs, while retaining the anonymity of their teachers and schools. We have developed and shared prototype reports and are in the process of gathering and acting on feedback to ensure they are useful. We have identified a tension: that MATs will often want to obtain more information than we can give them, to act on the potential insights in these reports. We consider this an opportunity to learn which additional data sources might be helpful to include in the TED in future. For example, MATs often want to know the classroom practices of their most effective teachers. Instead of identifying those teachers so a MAT's teaching and learning team can observe them, we are instead building a database of lesson transcripts and codifying the teaching practices evident within them. This will allow us in future to explore which teaching practices are associated with a higher TVA, which can answer the questions of interest to the MATs (and the sector in general), while retaining the anonymity of individual teachers and schools. The same approach may be used for further questions in the future.

This challenge relates to a point raised by our Strategic Advisory Board: as interest from MATs in our value-added analysis grows, so does the need to be clear about what these measures can and cannot support. We are therefore pairing the reports with explicit guidance on their limitations, emphasising triangulation and caution, and reinforcing that they are developmental tools rather than instruments of performance management.

Taken together, these steps move the work from establishing that teaching impact can indeed be estimated, towards making those estimates progressively more precise, more sustainable to produce, and more useful to the trusts and, ultimately, the teachers and pupils they are intended to benefit.

2.
What factors are associated with teaching impact?

Estimating teaching impact is only a first step. The larger purpose is to understand which factors are associated with it, so that recruitment, initial teacher training, and CPD can be directed towards what genuinely helps pupils. This section brings together the research questions that pursue that aim, spanning teachers' training and development, the schools and classes they work in, and their observed classroom practice. Because no single dataset can answer these questions, we combine schools' assessment data with HR and training records, staff surveys, and lesson-based evidence.

2.1 What do we want to know?

Our questions here are largely associational. As in Section 1, we aim to identify statistically significant patterns rather than to establish cause and effect. Where we find a factor associated with teaching impact, we share it as an association that can inform trusts' own operational decisions and to generate hypotheses to test in causal trials, such as those described in Section 3, rather than as evidence of cause and effect.

2a. Training and development pathways

We want to know which routes into teaching are associated with higher teaching impact. Our interest is in route types, for example a Postgraduate Certificate in Education (PGCE) against School Direct, rather than in individual providers, so that no training institution can be identified. The association between training route and teaching quality has been examined in the United States (Bardelli et al., 2023), but not in England. As we explain in the next section, the HR data available to us have so far been too incomplete to support this comparison, and we have had to adapt our approach.

2b. Teaching impact over time and retention

We want to know how teaching impact changes over a career, and how it relates to whether teachers stay in their school, their trust or the profession. Research from the United States has documented relationships between teaching quality and retention (Grissom and Bartanen, 2019), but these are not well understood in England. A central question is the trajectory of growth in teaching impact as experience increases. For example, a teacher's effectiveness might increase the same amount each year, or more quickly in the early years, with smaller gains continuing later in a career (Podolsky et al., 2019). Matthew Wiswall (2013) found that it is perhaps the case that teachers continue to improve after five years, but previous methodologies may have understated this growth

by aggregating all future years as one period, diluting increases in impact in that period. Additionally, Wiswall showed previous estimates of growth in teacher impact might have been biased by non-random attrition, i.e. average teacher impact might increase over time purely because less effective teachers leave the profession, rather than specific teachers becoming more impactful over time. We hope to bring some clarity to these questions by exploiting the longitudinal nature of the TED.

2c. School practices and characteristics

We want to know which features of a school are associated with its teachers becoming more effective. US evidence shows that teachers in schools with stronger professional environments improve faster (Kraft and Papay, 2014), and reviews of professional development emphasise the importance of school culture (Sims et al., 2021). We are interested in aspects of the working environment, such as leadership and collegiality, measured through validated survey scales (Sims, 2021), alongside concrete organisational practices. These include how pupils and teachers are assigned to classes, such as 'looping', where the same teacher is deliberately kept with the same class of pupils from one year to the next, on the basis that knowing the pupils well should help their learning.

2d. Observed teaching practices

We want to know which specific, observable classroom practices distinguish higher-impact teaching. Building on leading studies in the United States (Kane and Staiger, 2012) and England (Burgess et al., 2022), we aim to map the varied lesson-observation frameworks used across our trusts onto the wider literature, and to explore which practices are associated with teaching impact, and whether those associations vary with a teacher's experience, training or school.

2e. Analysis of Teaching Practice, using AI

We want to know whether AI tools can help us capture teaching practice at scale overcoming the cost and subjectivity that limit conventional lesson observation (Kane and Staiger, 2012). Through our ATP work, and building on approaches that apply text-as-data methods to lesson transcripts (Liu and Cohen, 2021), we set out to test whether transcripts of lessons can be reliably coded against a framework of teaching practices, and then related to teaching impact. The framework is aligned with the Initial Teacher Training and Early Career Framework (ITTECF), the Department for Education's statutory framework setting out what all trainee and early career teachers are entitled to learn. Coding lessons against practices highlighted in the ITTECF lets us ask which of those nationally defined practices are most strongly associated with pupil progress.

2.2 What have we learned so far?

Progress across the research questions in this section depends heavily on data collection, which has been the project's most persistent challenge. We now hold linked pupil and teacher data from seven trusts, at varying stages of completeness. This is enough to have begun work on all questions in this section, though several remain at the stage of assembling and refining data rather than producing findings. Two have already yielded early results.

2a. Training and development pathways

Our original plan to compare teaching impact by training route has not been possible, because the HR data received from trusts record training route inconsistently and are largely incomplete. We have therefore revised the question. Where we can obtain better HR data, we will still compare route types such as PGCE against School Direct. We will, however, be drawing on NIoT's own records of the teachers who trained through our programmes, including their programme assessments such as written assignments, and examining whether these are associated with later teaching impact. Linkage between our programme records and teachers in the TED runs through the unique Teacher Reference Number.

2b. Teaching impact over time and retention

This work is at an earlier stage. It depends on several years of linked data per teacher, which the TED is now accumulating, and on the value-added estimates becoming more precise as further data arrive. We are not yet able to report findings on how teaching impact evolves over a career or relates to retention. A significant constraint is the quality of HR data on teacher experience, since the years of continuous service are currently patchy across trusts.

2c. School practices and characteristics

This research question has two parts: school culture, and the allocation of classes to teachers.

On school culture, an earlier attempt at a standalone survey with one trust received very low response rates, so we changed the approach and embedded our questions in an existing survey that staff already complete. We added a short scale measuring the social capital between teachers to The Engagement Platform (TEP), the staff survey completed across participating trusts by approximately 8,000 teachers across 100 schools with a response rate of over 70%. These data have now been gathered and analysis is underway.

On the allocation of classes to teachers, John Jerrim, Becky Allen and Sam Sims produced the most comprehensive analysis of within-year class moves in England to date, using TED data from one trust covering 21 secondary schools and more than 10,000 pupils (Jerrim, Allen and Sims, 2026). Moving up a set was not followed by improved attainment in maths, English, history or geography, though it was in science. Moving down a set was followed by a fall in attainment in every subject analysed except geography. The tentative implication, which rests on observational data from a single trust and one year, is that the disruption of a mid-year move rarely repays itself. We will repeat this analysis with a further year of data, and extend it to compare within-year moves against changes made at the end of the year. The **pre-print of the paper** is available online.

We have also used the TED to answer a policy question posed by the Department for Education, on whether pupils with SEND are taught by less effective teachers. In one MAT comprising over 20 secondary schools and 17,000 pupils, we found that SEND pupils are not systematically allocated less effective teachers, and that teachers who are effective for non-SEND pupils are also effective for SEND pupils. The **full analysis** has been published on the NIoT website.

2d. Observed teaching practices

We have obtained lesson-observation data from one trust which allows us to explore which recorded practices are associated with teaching impact. We are currently running our value-added models on the data, after which we will code the observations and test their associations with impact. Linking observed practice to teaching impact is analytically demanding, and standard approaches have often produced weak results in the wider literature. We are therefore exploring a range of modelling methods to identify the most effective way of detecting these associations.

2e. Analysis of Teaching Practice using AI

Observation and structured feedback are among the most powerful levers for developing early career teachers, yet conventional mentoring is limited by the mentor needing to be physically present, by the partial nature of what any observer can take in, and by the time it demands. Lesson recordings and transcripts, by contrast, contain rich and recoverable signals of teaching practices, and recent research shows that AI methods can identify some teaching practices from transcripts with accuracy comparable to human raters (Xu et al., 2024), and that automated feedback can change teachers' practice (Demszky et al., 2024). The ATP work is exploring the opportunities and risks of doing this at scale: a secure recording tool, a custom AI model trained on expert-labelled transcripts to identify ITTECF-aligned practices, and a mentor-facing feedback interface. Considerable work has been conducted into the legal and ethical groundwork for recording lessons, which we describe in Section 4. With that in place, we have completed an initial pilot in which teacher development experts assessed lesson transcripts and recordings against the first version of our teacher-practice codebook, generating the labelled data needed to refine the framework and, in time, to train the model. The pilot report is being finalised and is expected to be published in late 2026.

2.3 What happens next?

Over the coming year the emphasis will shift from assembling data towards reporting associations between teaching impact and teachers' characteristics and practices. The following section sets out when we expect each to become available.

2a. Training and development pathways

Our next step is to link NIoT records of the teachers who trained on our programmes to their later teaching impact, through the Teacher Reference Number, captured as trusts join the TED. This depends on updated data-sharing agreements and on sufficient numbers of teachers being matched across trusts. We expect this analysis to become available during 2027.

2b. Teaching impact over time and retention

Having established the model with the two years of data we now hold, we expect to produce initial findings by the end of 2027, and to update them in 2028 as a further tranche of data arrives. The main constraint is the quality of HR data on teacher experience. At present the date a teacher qualified is patchy as many schools only ask to see proof of qualification, rather than recording specific details. This means that for many teachers the only reliable marker we have is when they entered the TED. Improving these data is a priority, as it will let us distinguish experience from the effect of the training a teacher received.

2c. School practices and characteristics

On school culture, our next step is to analyse the social-capital data now gathered through The Engagement Platform and link it to teaching impact at school and department level. We expect initial findings on the relationship between a school's professional environment and teacher effectiveness by the end of 2026 or early 2027. On the allocation of classes to teachers, and building on the within-year class-moves analysis, we will extend the work using timetable data now held from two trusts, which records more precisely who teaches whom and when. We expect to share further outputs in early 2027.

2d. Observed teaching practices

This analysis is underway. Having obtained lesson-observation data from one trust, we will code those observations and test which practices are associated with teaching impact. We expect initial findings by the end of 2026 or very early 2027.

2e. Analysis of Teaching Practice using AI

The pilot report will be published in late 2026. Over the 2026–27 school year we will collect a larger body of labelled lesson transcripts as the recording tool scales up, and then train the AI model. The model will then be tested in a randomised controlled trial of AI-supported mentoring for early career teachers during the 2027–28 school year, with analysis expected by the end of 2028.

3.
Can we improve teaching impact?

This section looks at how TIDE aims to turn evidence into practical improvement, by examining the tools, feedback and trial designs through which insights about teaching impact might be used to help teachers develop, and ultimately benefit pupils.

3.1 What do we want to know?

Estimating teaching impact – and understanding what is associated with it – matters only if that knowledge can then be turned into practical improvement. We want to take learnings from the first stages of the project and use them to help teachers become more effective, and in turn to improve outcomes for pupils. The question is whether the evidence TIDE generates can be translated into useful, usable interventions, and whether those interventions work.

Two areas of teacher development offer a natural starting point. Both are widespread, resource-intensive, and currently rest on limited evidence about their effectiveness. The first is the mentoring of early career teachers. This sits at the heart of the Early Career Teacher programme, yet its effectiveness in England is not well understood. The second is peer observation. Most teachers experience it routinely, but it is rarely informed by evidence on which observed practices matter most.

Beyond these specific interventions, we have a wider aim. We want to demonstrate how the TED and its infrastructure can be used to develop, implement and rigorously evaluate improvements to teacher development, quickly and at a lower cost. That way, we, and others, can repeat the process across many more questions in future.

3.2 What have we learned so far?

We do not yet have direct evidence on whether TIDE-informed interventions improve teaching impact. The trials that will test this are not due to run until the 2027–28 school year. Much of the work so far has been the groundwork needed to design interventions that stand a real chance of being useful once they are trialled.

The first strand of that groundwork is a detailed understanding of our partner trusts. We have learned what data each trust holds and how reliable it is, the shape of their assessment systems, where responsibility for teacher development sits, and how they understand their own recruitment, retention and development challenges. This knowledge is what will allow any future intervention to be designed around how trusts actually work in practice.

The second is the Analysis of Teaching Practice work described in Section 2. The recording tool developed there is likely to be the main mechanism through which any future intervention is delivered. Building it is therefore also laying the foundation for those interventions. It has meant exploring how mentors could use AI-supported feedback, how this might relate in detail to the ITTECF and ECT programmes, and what the experience would be for early career teachers. It has also involved establishing whether a tool can be developed robustly and securely enough to roll out at scale. Its design has been shaped throughout by those who will use it. Our Expert Working Group (see Appendices), for example, steered the work towards providing AI feedback to mentors rather than directly to teachers, and towards a developmental framing rather than one of judgement.

The third and most immediate output is our trust-level MAT reports. These present each trust with an interpretable picture of teaching impact across its schools, without identifying any individual teacher. Each report includes the distribution of teaching effectiveness within a school. This is shown using the anonymised format illustrated earlier in this report, where each teacher appears as a single point and no individual can be identified. Where a school is small enough that showing this distribution might make individuals identifiable, it is not shown. The reports also show how effective teachers are distributed across pupils with different characteristics, such as special educational needs, gender, English as an additional language, and pupil premium eligibility. In addition, each report includes a detailed audit of a trust's assessments. This identifies assessments that may be too easy, too difficult, or poorly correlated with other assessments. It also examines how well each assessment predicts subsequent GCSE grades. That last point helps trusts understand whether each round of assessment generates genuinely new information or largely repeats what earlier assessments already showed. This can, in turn, help them reduce unnecessary assessment and the workload it creates. We have shared prototype reports with our founding MATs and are working with them, through a dedicated workshop, to understand how the reports are read and how they can best support conversations about teacher development.

3.3 What happens next?

The next phase moves from preparation to testing whether TIDE's outputs can support real improvement in practice. In the near term, we will continue to develop the MAT reports alongside the trusts that use them. We will refine how teaching-impact data are presented and learn what makes it genuinely useful. This work is valuable in itself. It also serves as a testing ground for the larger question of how evidence about impact can be turned into action.

The more significant step is the move towards trialling. Over the 2027–28 school year we will run a randomised controlled trial. It will test whether AI-supported mentor feedback changes early career teachers' practice and, in turn, improves their pupils' attainment. We will carefully monitor and evaluate any unintended consequences of AI-supported mentoring, such as a change to the ECT mentor relationship or ECT's experiences of feedback. We will publish the trial protocol on our website in advance. We are also considering how a second strand of work, focused on peer observation among more experienced teachers, might best be developed. We expect to set out our approach early in the new year.

Beyond these specific trials, our findings from across the project will be shared widely. We want them to inform teacher development throughout the profession, rather than only in the trusts we work with directly. We will tailor how we do this to four audiences.

For school and trust leaders, we will share findings with our founding MATs, with other trusts contributing data to the TED, and with the NIoT's network of around 250 associate college schools. We will do this through tailored reports, webinars and workshops. We will also present our findings at ResearchED and the Confederation of School Trusts (CST) Conference.

For teacher educators, we will present at sector conferences and produce accessible practitioner resources. We will also integrate findings into the NIoT's own programmes, which reach teachers at every stage from initial training, National Professional Qualifications, to trust leadership. The NIoT's own Evidence Portal, which exists to promote well-evidenced approaches to teacher development, gives us a further route to reach the sector.

For policymakers, we will seek to inform the Department for Education's guidance on initial teacher training and continuing professional development. We will work in partnership with other training providers, so that the evidence, rather than any single organisation, drives the case for change.

For the research community, we will submit our findings for publication in the most impactful journals possible. We will present at academic conferences such as BERA and NASBTT. We will also provide access to the TED so that other researchers can build on our work.

Where appropriate, we will seek to secure coverage in the education press.

Finally, where an association looks important, we will look for ways to act on it and to evaluate whether doing so helps. For example, if we were to find that stronger social capital between teachers was associated with faster growth in teaching impact, we would work with partner trusts to find ways of strengthening it, and evaluate the result using causal methods such as randomised or quasi-experimental trials. If particular timetabling arrangements were found to support the development of early career teachers, we would look to introduce and test them in the same way. In each case the evaluation will use our teacher value-added measure and the TED. This demonstrates again how the infrastructure can turn a research insight into a tested improvement.

4.
Responsible use of data and AI

TIDE works with sensitive information. It links pupils' attainment to the teachers who taught them, and, for the ATP project, uses lesson transcripts from recordings of real lessons. This section sets out how we protect data and identity, and how we govern the risks of recording lessons.

4.1 The Teacher Education Dataset (TED)

The TED brings together data that MATs already hold so that researchers can learn how to support school improvement and teacher development without any individual school, teacher or pupil being identifiable. The estimates we produce are for understanding patterns across the system, not for judging individuals. Our teacher-impact measures are sector-level developmental and research tools, not instruments of individual performance management, appraisal or pay. Therefore, manually extracted data are carefully pseudonymised before being made available to the TIDE research team. However, our longer-term goal is for other researchers to be able to access the data safely and securely, which means we must build an infrastructure for them to do so.

4.2 OpenSAFELY and NHS health data

The TED's planned researcher-facing infrastructure is modelled on OpenSAFELY, a platform first built for the NHS. Its story shows both what secure, large-scale data analysis can achieve and why it must be done carefully.

OpenSAFELY was created in the first months of the COVID-19 pandemic in 2020. Policymakers urgently needed to know who was most at risk of dying from the virus, but the answer lay in GP records, some of the most sensitive data the state holds. Providing that data to researchers while adhering to data protection laws would have been slow, and might have created serious privacy risks. The team at the Bennett Institute at the University of Oxford, working with the NHS and others, therefore took a different approach. Instead of providing researchers with raw data, they created an interface where researchers can first write and test their code against representative dummy data. Researchers can then submit their query to analyse the real data that never leaves a secure environment. Every line of analytic code is published openly for inspection and reuse. The team built a working platform in around six weeks, and its first study, covering the linked records of 17 million adults, produced some of the earliest robust evidence on

COVID-19 risk factors and directly shaped national shielding policy.

Since then, the platform has grown to cover the pseudonymised GP records of around 58 million patients, close to the entire population of England. It has supported a large body of independent research. The approach has earned the trust of the GP profession, patient groups and privacy campaigners. Having proved itself during the pandemic, the platform is now opening-up to health researchers more broadly.

4.3 OpenSAFELY-Schools and the TED

Education data brings its own sensitivities, and specifically, much of the data concerns children. That raises not only the data protection questions that any personal data brings under UK GDPR, but also safeguarding questions about how children's information is held and used. This requires a partner with a proven, tried-and-tested approach to sensitive data at scale. Our long-term plan is therefore for the TED to be accessed by researchers via OpenSAFELY-Schools, built by the Bennett Institute. As described above, researchers will never have access to the TED's raw data. Researchers write analyses that run against the data remotely, and all analytic code will be published.

We are scaling this work up from 2026. We are recruiting more MATs and working with management information system (MIS) providers so that we can import their data structures efficiently and with appropriate permissions. From 2027, we will look to make the TED available to approved researchers on application, in the same way that OpenSAFELY makes health data available. All data queries and research projects will be logged and stored openly on GitHub, so that analyses can be reproduced and reviewed. We also intend to build a set of reusable code packages, so that researchers can quickly apply methods we have developed, such as our teacher value-added measure, and adapt the underlying code if they wish.

Our longer-term ambition is for the TED to outlast TIDE. Once this project completes, the dataset and its infrastructure will remain available to researchers under these conditions. We will promote it to the research community internationally, and to the organisations that fund and evaluate interventions. The aim is to increase the speed at which rigorous education research can be done, while maintaining the highest standards of data protection, and to build on the exploratory groundwork set out in this report.

4.4 Ethical and lawful recording of lessons

The ATP project required the most intensive governance work, because recording lessons involves the personal data of teachers and, incidentally, of pupils. Over the past year, and in consultation with lawyers, data protection specialists, officials from the Department for Education and the Children's Commissioner's office, we completed a Legitimate Interests Assessment (LIA), a Data Protection Impact Assessment (DPIA), and a data-sharing agreement (DSA) with participating schools. These were reviewed through the NIoT's Ethics Committee and by our Data Protection Officer (DPO).

This work established distinct lawful bases for the project's two purposes. Supporting ECTs' development rests on Public Task while developing and evaluating the AI model rests on Legitimate Interests. Several safeguards are built into the design. Recording is voluntary and teacher-initiated, rather than continuous. Devices are positioned to focus on the teacher and to avoid capturing pupils on video. Recordings are stored securely within NIoT systems; access to the tool is protected by two-factor authentication; and each recording can only be viewed by the ECT or mentor who made it, or by an ECT's mentor if the ECT chooses to share it. Recordings are automatically deleted after 30 days, and teachers can delete their own recordings at any time. Teachers can also opt out of their lesson transcripts being used for research each time they start a new recording. Any research is conducted on redacted transcripts from which names and other identifying details have been removed. The AI's outputs will be advisory only, and their use for capability, appraisal or disciplinary purposes is contractually prohibited.

Before any recording takes place, schools are asked to share a notification with parents and age-appropriate information sheets with pupils. Teachers are asked to let pupils know whenever they are recording, and to position the device to focus on themselves and avoid capturing pupils' faces. If something sensitive is captured, such as confidential information or pupils coming into view, the teacher can delete the recording immediately after they stop recording, before it is uploaded. Parents can ask the school for their child's class to be excluded from research. The school passes this on to the relevant ECT or mentor, who then selects the research opt-out each time they record that class, so that transcripts from those lessons are not used for research. Because a lesson recording cannot selectively exclude one child, a single opt-out removes the whole class from research. An opt-out does not stop the lesson being recorded for the ECT's own development, although some schools may choose not to record opted-out classes at all.

As we develop this work, we will evaluate not only the potential benefits of AI-supported mentoring, but also its risks and any unintended consequences, and we will report on both.

4.5 Responsible interpretation

Responsible use extends to how findings are read. Estimates of teaching impact are imprecise for any individual teacher, with wide confidence intervals. There are two main reasons for this. The first is the within-teacher variation shown in Figure 7, where the same teacher's estimated impact varies considerably across the different classes they teach. The second is random variation in assessment results. This can reflect how a group of pupils happened to perform on the day, noise arising from assessments of variable quality, or unobserved characteristics of pupils that lie outside our data. All of these widen the confidence intervals around an individual estimate. We are working to narrow them, by collecting better-quality data on pupils and by accumulating several years of results for each teacher. For now, at the 95% confidence level, these intervals remain wide. We are explicit about this wherever the measures are presented, particularly as interest grows among trusts in using the analysis themselves. We pair our trust-level reports with clear guidance on their limitations. Genuine patterns should not be over-interpreted, and noisy individual estimates should not be mistaken for firm judgements.

5.
Case Study – Oasis Community Learning

Sara-Jayne Martin, Oasis' National Lead for Professional Learning, spoke to Raj Chande about why her trust is participating in TIDE, and their experience of working with the team so far.

Why did the trust decide to take part in TIDE?

We really care about professional development. Experience matters, but we also don't want to set our priorities just because they feel right or because someone personally connected to them. It would be immensely valuable to have an objective starting point, grounded in our own data, for how we support and develop teachers. It's easy in education to get pulled towards what we individually or emotionally connect to, and this gives us something more objective to build those decisions on.

What are you hoping TIDE will achieve?

The hope is that it lets us be much more precise about how we train and support teachers, and about the development decisions we make as a trust. Teachers often feel their development time is spent on things they don't think matter most, so being more focused on what genuinely helps them improve is an exciting prospect.

When you saw your trust's data in your bespoke MAT Report, what stood out?

The clearest thing was that even the most effective teachers show significant variance across the classes they teach. We don't acknowledge enough as a profession that noone is perfect at this. You can be a highly effective teacher and still find a particular group hard. Seeing that so clearly in the data helps depersonalise development. I talk about it all the time now when I develop leaders: it's not a judgement about you as an individual, it's about incremental improvement over time.

Has it changed how you think about your assessment data?

The section on the quality of the assessments was a really interesting provocation for me about how we interrogate the validity and reliability of our assessments, rather than carrying them out and just assuming the data is sound, and how we help teachers and leaders think that way too.

In your experience, what makes the difference to how effective a teacher becomes?

It's a combination of three things working together. First, a school culture of continuous development, where observation serves growth rather than judgement. Where that's missing, there's a real ceiling on how effective a teacher can become. Then the precision of the training and mentoring, well-diagnosed and matched to the teacher's need at that moment, and the clarity with which they understand the change they're trying to make, ideally with a clear model of what they're aiming for.

How does the trust decide which teachers are assigned to which classes and sets?

At the moment there isn't a unified position. Schools are largely left to allocate their staffing as they see fit. Historically the most effective teachers are pushed towards Year 11 at the expense of Key Stage 3, and that's probably been true across the trust; but as we start to think seriously about developing expertise over time, this is something we need to take a clearer position on. One of our schools has just moved to linear timetabling so every teacher takes Year 7 to 11, which sends a powerful message that every moment in a child's education matters.

What do you make of the finding on mid-year set changes?

Intuitively it resonates: any change to a child's education carries a lag, and a mid-year move asks the teacher to build a new relationship and learn a child's history part-way through the year. We don't have a trust-wide strategy on this, but a piece of research like that might rightly orient people's thinking towards it.

Can TIDE help you increase the impact of your teachers?

Yes. First of all, our MAT report gives us an objective starting point to exemplify that everyone is strong at some things and less strong at others, which lets us be objective about development, and that's exactly where teacher development tends to fall down, when it starts to feel judgemental. And because it's our data and our context, it removes the "that was done somewhere else, it's not relevant to us" reaction. Above all it opens up conversation about our teaching and development priorities, which is powerful in itself.

How does the trust use evidence and data to shape CPD now?

We draw on the core evidence base the sector has normed on, such as the **EEF toolkit** and the **NIoT Evidence Portal**. Each strand of CPD then uses assessment cycles, school visits and qualitative insight to shape what's delivered. We've been on a journey to bring greater coherence across those strands and a more unified understanding of what effective development looks like. The most valuable thing this project could do is help us join that data-and-evidence thread up consistently across all our schools.
This is something we're actively refining, as I suspect most large trusts are.

How do you feel about using video and AI to develop teaching practice?

Across our NPQs and leadership development we're actively trying to normalise analysing teaching through video. It's an incredibly useful tool for reflection, coaching and seeing progress over time, and there's senior agreement that we want people using it. That said, teachers and leaders can be nervous about being filmed, and where leaders are reluctant to model it, teachers are reluctant to take part. People become comfortable when a leader has gone first and shown it's safe, or when it quickly helps them improve something in their own practice.

How important is it that the data used in TIDE stays anonymised and away from performance management?

It's crucial that people know their name won't appear on a scale or a list. This work is about using a large dataset to inform teaching and development priorities, not to penalise anyone. There's a natural nervousness about the data being used for judgement, but the way the intention, the security and the handling of the dataset have been explained has left me and colleagues feeling genuinely confident about it.

Have you felt reassured about the data protection behind the lesson recording in the ATP project?

I'm fairly relaxed about it. We've entrusted our wider data to the project already, and there's no reason not to trust the assurances and safeguards being given. We've been through the agreement carefully and our data protection officer has gone through it with a fine-tooth comb, which reassures me at a central level. How it lands at school level is the part I'm most curious to see as we go.

6.
What next for TIDE?

This final section looks ahead across the whole programme. While the earlier sections set out what comes next for each research question in detail, here we step back and describe the overall direction of travel, first for the coming year and then for the remainder of the project.

The first phase of TIDE has largely been about building foundations. We have assembled the datasets, partnerships, infrastructure and governance that make ambitious research possible. With those in place, the programme now moves from establishing that teaching impact can be measured towards understanding what is associated with it, and then towards testing whether that knowledge can be used to improve teaching.

The next 12 months

Over the coming year the emphasis shifts from assembling data to reporting findings. As further tranches of data arrive and our estimates become more precise, the analyses described earlier will begin to produce results. This means more outputs across the programme's research questions, published as they are ready, rather than a single report at the end. By the time of our next annual report, we expect to have shared initial findings on several of the questions set out in Section 2, alongside continued refinement of our measures of teaching impact.

The infrastructure will also grow over this period. As set out in Section 4, we are scaling up the TED from 2026. We are recruiting more trusts and working with management information system providers to bring their data in efficiently and securely. This expansion will allow the analyses to become more precise and to extend across a wider range of schools.

The remainder of the project

Beyond the coming year, the work moves towards causal tests. The associations we identify suggest what the drivers of teaching impact might be. Over the 2027–28 school year we will begin to verify these through trials, starting with a randomised controlled trial of AI-supported mentoring for early career teachers, and looking to extend this approach to other promising findings. This is how the project moves from describing what is associated with effective teaching, to testing what actually improves it.

The infrastructure ambition also extends beyond TIDE itself. From 2027 we will look to open the TED, through OpenSAFELY-Schools, to approved researchers on application. Our aim is that the dataset outlasts the project, so that many more researchers can build on the exploratory work begun here, and education research as a whole can be done faster while maintaining the highest standards of data protection.

Get involved

The value of TIDE grows with every trust that takes part. If you lead or work in a multi-academy trust and would be interested in contributing your data to the TED, we would be glad to talk. The more trusts involved, the more the whole sector can learn about how to develop teachers effectively.

To discuss contributing data, please contact us at **ted@niot.org.uk**

We would also welcome contact from researchers who would like to use the TED to answer their own questions, and from teacher development experts and school leaders who would like to help shape the work.

For research collaboration or any other enquiry, please contact **research@niot.org.uk**

Bibliography

Allen, R., Jerrim, J., Sims, S., Chande, R., Coe, R., Davey, C., Maud, C. and Ventista, O. (2026) 'Challenges and validation in estimating teacher value-added in English secondary schools', International Journal of Research & Method in Education, 49(4), pp. 431–451. Available at: **https://doi.org/10.1080/1743727X.2026.2687439.**

Bardelli, E., Ronfeldt, M. and Papay, J.P. (2023) 'Teacher preparation programs and graduates' growth in instructional effectiveness', American Educational Research Journal, 60(1), pp. 183–216.
Available at: **https://doi.org/10.3102/00028312211066867.**

Burgess, S., Rawal, S. and Taylor, E.S. (2022) Characterising effective teaching. London: Nuffield Foundation.
Available at: **https://www.nuffieldfoundation.org/wp-content/uploads/2022/05/Burgess-Characterising-Effective-Teaching-Full-Report-April-2022.pdf** (Accessed: 20 October 2026).

Chetty, R., Friedman, J.N. and Rockoff, J.E. (2014) 'Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates', American Economic Review, 104(9), pp. 2593–2632.
Available at: **https://doi.org/10.1257/aer.104.9.2593.**

Coe, R., Ventista, O., Chande, R., Maud, C., Kime, S. and Dillon, S. (2024) Estimating teacher value-added from schools' internal assessments in England. OSF Preprints.
Available at: **https://doi.org/10.31219/osf.io/tqnjk.**

Demszky, D., Liu, J., Hill, H.C., Jurafsky, D. and Piech, C. (2024) 'Can automated feedback improve teachers' uptake of student ideas? Evidence from a randomized controlled trial in a large-scale online course', Educational Evaluation and Policy Analysis, 46(3), pp. 483–505.
Available at: **https://doi.org/10.3102/01623737231169270.**

Education Policy Institute (EPI) (2021) The cost of high-quality professional development for teachers in England. Authored by J. Van den Brande and J. Zuccollo. London: Education Policy Institute.
Available at: **https://epi.org.uk/wp-content/uploads/2021/07/2021-Cost-of-quality-teacher-cpd_EPI.pdf (Accessed: 20 October 2026).**

Grissom, J.A. and Bartanen, B. (2019) 'Strategic retention: Principal effectiveness and teacher turnover in multiple-measure teacher evaluation systems', American Educational Research Journal, 56(2), pp. 514–555.
Available at: **https://doi.org/10.3102/0002831218797931.**

Hanushek, E.A. and Rivkin, S.G. (2010) 'Generalizations about using value-added measures of teacher quality', American Economic Review, 100(2), pp. 267–271. Available at: **https://doi.org/10.1257/aer.100.2.267.**

Jerrim, J., Allen, R., Palma Carvajal, M., Chande, R., Coe, R., Davey, C., Dillon, S., Maud, C., Sims, S. and Ventista, O. (2025) 'Initial estimates of teacher value-added in English primary schools', British Educational Research Journal, pp. 2942–2963. Available at: **https://doi.org/10.1002/berj.4207**

Jerrim, J., Allen, R. and Sims, S. (2026) Should pupils move class during the academic year? Evidence from a multi-academy trust in England. Working paper. Zenodo. Available at: **https://doi.org/10.5281/zenodo.23082706.**

Kane, T.J. and Staiger, D.O. (2008) Estimating teacher impacts on student achievement: An experimental evaluation. NBER Working Paper No. 14607. Cambridge, MA: National Bureau of Economic Research.
Available at: **https://www.nber.org/papers/w14607** (Accessed: 20 October 2026).

Kane, T.J. and Staiger, D.O. (2012) Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains. Seattle, WA: Bill & Melinda Gates Foundation.
Available at: **https://files.eric.ed.gov/fulltext/ED540960.pdf**
(Accessed: 20 October 2026).

Kane, T.J., Taylor, E.S., Tyler, J.H. and Wooten, A.L. (2011) 'Identifying effective classroom practices using student achievement data', Journal of Human Resources, 46(3), pp. 587–613.
Available at: **https://doi.org/10.3368/jhr.46.3.587.**

Koedel, C., Mihaly, K. and Rockoff, J.E. (2015) 'Value-added modeling: A review', Economics of Education Review, 47, pp. 180–195.
Available at: **https://doi.org/10.1016/j.econedurev.2015.01.006.**

Kraft, M.A. and Papay, J.P. (2014) 'Can professional environments in schools promote teacher development? Explaining heterogeneity in returns to teaching experience', Educational Evaluation and Policy Analysis, 36(4), pp. 476–500.
Available at: **https://doi.org/10.3102/0162373713519496.**

Liu, J. and Cohen, J. (2021) 'Measuring teaching practices at scale: A novel application of text-as-data methods', Educational Evaluation and Policy Analysis, 43(4), pp. 587–614.
Available at: **https://doi.org/10.3102/01623737211009267.**

Podolsky, A., Kini, T. and Darling-Hammond, L. (2019) 'Does teaching experience increase teacher effectiveness? A review of US research', Journal of Professional Capital and Community, 4(4), pp. 286–308.
Available at: **https://doi.org/10.1108/JPCC-12-2018-0032.**

Sims, S. (2021) 'Why do some schools struggle to retain staff? Development and validation of the Teachers' Working Environment Scale (TWES)', Review of Education, 9(3), e3304.
Available at: **https://doi.org/10.1002/rev3.3304.**

Sims, S., Fletcher-Wood, H., O'Mara-Eves, A., Cottingham, S., Stansfield, C., Van Herwegen, J. and Anders, J. (2021) What are the characteristics of effective teacher professional development? A systematic review and meta-analysis. London: Education Endowment Foundation.
Available at: **https://educationendowmentfoundation.org.uk/education-evidence/evidence-reviews/teacher-professional-development-characteristics** (Accessed: 20 October 2026).

Wiswall, M. (2013) 'The dynamics of teacher quality', Journal of Public Economics, 100, pp. 61–78.
Available at: **https://doi.org/10.1016/j.jpubeco.2013.01.006.**

Xu, P., Liu, J., Jones, N., Cohen, J. and Ai, W. (2024) 'The promises and pitfalls of using language models to measure instruction quality in education', in Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers). Mexico City: Association for Computational Linguistics, pp. 4375–4389.
Available at: **https://doi.org/10.18653/v1/2024.naacl-long.246.**

Appendices

Glossary

Confidence interval: A range around an estimate that shows how much uncertainty surrounds it. A wide confidence interval means the true value could plausibly fall anywhere across a broad range, so we cannot be precise. In this report, most individual teachers' value-added estimates have confidence intervals that cross zero, which is why the measure cannot be used to judge an individual teacher.

Early Career Framework (ECF): The statutory framework in England setting out the training and support that teachers are entitled to in their first two years in the profession.

Free School Meals (FSM): A measure of pupil disadvantage, used as a control in our models. Pupils are eligible for free school meals if their family meets certain income thresholds.

Initial Teacher Training and Early Career Framework (ITTECF): The Department for Education's statutory framework setting out what all trainee and early career teachers are entitled to learn. In this project, lesson transcripts are coded against ITTECF-aligned teaching practices.

Multi-Academy Trust (MAT): A group of schools run under a single trust. Because trusts often use common, internally standardised assessments across their schools, they make it possible to compare teaching impact at scale.

OpenSAFELY-Schools: The secure data infrastructure being built for this project, modelled on the OpenSAFELY platform used for NHS health data. Data never leaves the trust's control, researchers run analyses against it remotely, and all analytic code is published, so the data itself is never exposed.

Pseudonymisation: Replacing identifying information, such as names, with codes, so that individuals cannot be directly identified from the data. In this project, trusts pseudonymise data before it reaches the research team.

Teacher Education Dataset (TED): The dataset at the heart of TIDE. It links pupils' assessment results to the teachers who taught them, alongside pupil and teacher characteristics, drawn from participating multi-academy trusts.

Teacher Reference Number (TRN): A unique identifier held for each teacher in England, used in this project to link NIoT's own records of programme participation to teachers in the TED.

Teacher value-added (TVA): Our measure of a teacher's estimated contribution to their pupils' progress, separate from other influences on attainment such as prior attainment and pupil background. It is the technical approach used to estimate what this report calls a teacher's impact.

Teaching and Learning Responsibility (TLR): A payment made to teachers who take on additional leadership or responsibility, recorded in HR data.

Value-added: A general approach to measuring impact by comparing actual outcomes against what would be expected given prior attainment and other characteristics. Teacher value-added applies this idea to estimate the impact of individual teachers.

Our Strategic Advisory Board (SAB)

We have established a Strategic Advisory Board (SAB) to offer independent advice, support, and critical evaluation for the duration of the project.

The board is chaired by Professor Sir Chris Husbands and members include policymakers and policy influencers, unions, teacher educators, research and technical experts.

Emma Hollis	CEO, National Association of School-Based Teacher Trainers (NASBTT)
Steve Rollett	Deputy Chief Executive, Confederation of School Trusts (CST)
Niamh Sweeney	Deputy Chief Executive, Confederation of School Trusts (CST)
James Bowen	Director, NAHT
Professor Anna Vignoles	University of Cambridge
Professor Steve Higgins	Emeritus Professor, Durham University
Emily Yeomans	co-CEO, Education Endowment Foundation
Jack Worth	School Workforce Lead, NFER
Jon Andrews	Head of Analysis and Director for School System and Performance, Education Policy Institute
Dr Julie Belanger	Director, Better Purpose
Dr Rolf Strietholt	Co-Head of Research and Analysis, IEA
Jason Elsom	Chief Executive, Parentkind
Caroline Halls	Chief Economist & Head of School Strategy, Department for Education (DfE)
Juliette Cammaerts	Director of Policy, Planning and Performance, Children's Commissioner
Mark Leswell	School Improvement Research Lead, Swale Academies Trust
Amy Mitchell	Chief Impact Officer, Teach First
Jen Persson	Director, Defend Digital Me

Our Expert Working Group (EWG)

The EWG includes members from the TIDE research and NIoT programmes teams, as well as experienced teacher educators and development leads from diverse school contexts across England. Members have been selected for their experience, expertise, and willingness to take part. Members of the EWG will:

- Engage in meaningful collaboration with researchers and fellow educational leaders to bring forward the insights and challenges of today's classrooms.
- Impact teacher development by guiding the project to address key aspects of classroom practices and teacher professional development to improve pupil outcomes and aid retention.
- Be part of a supportive network of educators and researchers passionate about making a difference.

The EWG members are:	
Victoria Gardiner-Earl	Deputy Director, Newcastle Research School
Kate Hurley	Education Consultant, Polesand Education
Anupa Matharu	Assistant Headteacher, Firs Primary School, Washwood Academy Trust
Hannah Hamid	Regional Leader of Inclusion, Europe and US, Cognita Schools
Jenny Dixon	Deputy Director of Education: People Development, Aquinas Education Trust
Patricia Hogan-Hussein	Primary school teacher and Teachers' Coach, St Mary's RC Primary School
Emilio Bayarri-Torres	Secondary Development Lead, Teach First
Lucy Bomber	Consultant education advisor, Teach First and ImpactEd
Amy Carlile	Assistant Director of Education, LEO Academy Trust
Camron Mills	Headteacher, Firs Primary School, Washwood Academy Trust
Luke Jackson	Appropriate Body Lead, Star Teaching School Hub
David Aldridge	Dean of School of Curriculum and Childhood, Marino Institute of Education
Dan Smith	Associate Dean at School of Education, La Trobe University, Australia
Pip Sanderson	Director of Teacher Development, NIoT
Ellen Turner	Head of Evaluation, NIoT
Shaun Dillon	Director of Data, NIoT
Calum Davey	Executive Director of Research and Best Practice, NIoT
Raj Chande	Senior Research Associate, NIoT
Bronte McKeown	Senior Research Fellow, NIoT
Claire Maud	Research Fellow, NIoT


NATIONAL
INSTITUTE
of Teaching