2024

Primary Science and TIMSS

MARY OLIVER, MIKE ADKINS, ROSIE SMITH

Table of Contents

Acknowledgements

The Nuffield Foundation is an independent charitable trust with a mission to advance social well-being. It funds research that informs social policy, primarily in Education, Welfare and Justice. 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. The Foundation has funded this project, but the views expressed are those of the authors and not necessarily the Foundation. Visit www.nuffieldfoundation.org

Nuffield Foundation

Executive Summary

This study uses data from the 2019 Trends in International Mathematics and Science Study (TIMSS), to better understand factors impacting pupil achievement in science in Grade Four (Year Five in England). TIMSS is an international assessment project conducted by the International Association for the Evaluation of Educational Achievement (IEA) and is focused on Mathematics and Science conducted at the fourth and eighth grade level (Year 5 and 9 within England). In 2019, 64 countries took part, and the TIMSS curriculum model compares the intended, implemented and attained curriculum with questionnaires targeted at the pupil, home, teacher and school principal level. To ensure comparisons across different countries and jurisdictions, TIMSS has a clearly defined approach to sampling from the school population agreed between each national coordinator and TIMSS sampling experts.

Science assessment in TIMSS has been constructed on frameworks developed by the participating countries for each curriculum area and for each grade. The cognitive items aimed to assess pupils' achievement in the domains of knowing, reasoning and application (Martin et al., 2020). These are measured using a science test and estimated using a measurement model producing five plausible values, representing the range of pupil performance. A comparative secondary data analysis of pupils in primary schools provides an international perspective to identify the relative strengths of factors within pupils, schools and classrooms and determine associations between pupil achievement and interest in science. In this report, we also explore how teachers' professional development, years of teaching, and instructional practices are associated with achievement in science.

Teachers were asked in the teacher survey about the school's emphasis on academic success, school environment; pupils' behaviour, rules, safety, fairness; about being a teacher, including how they felt (meaning and purpose, enthusiasm) workload (teaching hours, pressure from parents, admin tasks, preparation for class and curriculum changes).

In England, 3,396 primary pupils from 139 primary schools participated, including eight independent schools and most being 'community' or state-funded schools. An in-depth examination was conducted with an extract of the TIMSS data for England linked to the National Pupil Database (NPD), held by the by the International Statistics division within the Department for Education (DfE). This allows for analysis of factors including free school meals (FSM), ethnicity as well as allow for a quasi-longitudinal study on pupil progress from the Foundation Stage Profile. The DfE were able to link 3319 TIMSS participating primary school pupils to their existing records which provided important measures of deprivation for the pupil (whether a pupil has been eligible for free school meals at certain time points – such as the in the last 3 or 6 years) whether pupils were ever recipients of the free school meal entitlement), and the Income Deprivation Affecting Children Index (IDACI) which is a postcode level measure of the proportion of children under the age of 16 that live in low income homes. Measures of prior achievement were more limited given the age of the pupils at the time of the 2019 TIMSS fieldwork, with follow on tests at Key Stage 2 (KS2) cancelled due to Covid. We used the recorded Phonics mark from the screening check in Year 1 to gain some sense of the level of their prior achievement.

The model (includes international data)

Our modelling proceeded on the basis of incremental improvements working from a basic null (or empty) model with no predictors, to a varying intercept model where indexes were added to account for clustering at the class, school and country level. We then increased the complexity of the model adding further pupil, teacher and school level variables and assessing model fit. Our outcome of interest was the TIMSS Grade Four (in England, Year 5) score which is centred on a score of 500 prior to including additional explanatory predictors.

The final model is built as follows.

- Firstly, we included predictors to account for pupil background which comprised of a binary sex of pupil variable (where the base case was female) and a continuous home resources for learning scale. In particular, the home resources scale asked pupils and parents on the number of books in the home, home study supports, the highest education level of either parent, and the highest occupational level of either parent.
- Secondly, we included three measures of pupil perceptions on whether they like learning science, their confidence in science and the instructional clarity in science lessons. In regards to instructional clarity, pupils were asked about aspects of their teachers' instruction: 'whether they knew what their teacher expects them to do, whether their teacher is easy to understand, has clear answers to their questions, is good at explaining things, and does a variety of things to help them learn … and explains a topic again when pupil do not understand' (TIMSS, p.475).
- We also controlled for the amount of instructional time received by the pupils which was mean-centred1, and whether they received any additional instruction in the form of further science lessons (such as with individual tutoring).
- Thirdly, at the teacher level we included 10 variables examining how science lessons are taught which had the following categories - "Every or almost every lesson" (base category), "About half the lessons", "Some lessons" and "Never". These variables examined pupils listening to the teacher explain; observe phenomena, the being demonstrated an experiment, planning experiments, conducting experiments, being presented data, interpreting data, using evidence, reading textbooks, memorising facts and doing fieldwork.
- Fourthly, we included a series of discrete variables (with three categories – a lot, some and none which was set as the base case) that measured the importance that teachers attach to specific assessment strategies. These were observing pupils as they work; asking pupils to answer questions during class; short, regular written assessments; longer tests; and long-term projects.
- Lastly, we also included a series of dichotomous variables to understand the impact of various professional development courses taken prior to the TIMSS survey. These included training on science content, science pedagogy, science curriculum, integrating technology, standard critical thinking, science assessment, past pupil needs, and integrating science subjects into the curriculum.

The model for pupils and teachers in England only

We also included:

- three demographic measures: sex of the pupil, the free school meal entitlement and IDACI score, and one measure of prior achievement, the score on the phonics screening check in year 1;
- three measures capturing the pupil experience of learning science: how clear their teaching has been, their enjoyment of science and their confidence in learning science;
- teachers' teaching strategies and experiences of CPD.

Findings

Teachers' professional development

The low level of participation by teachers in England limits generalisations from the analysis. Data show that teachers in England express a greater wish for continuing professional development in some areas, compared to others. Just over half the teachers completing the survey have taken part in CPD with a focus of the science curriculum. 72.4% of pupils were taught by a teacher who wishes to have future professional development opportunities in integrating

[1] Predictors were mean-centred to assist with our own model building to ensure that the intercept remained interpretable. We do however report the regression results as a coefficient plot without the intercept to simplify the results for the reader.

technology, and only 16.9% of pupils were taught by a teacher who had already had training in this area. Around 60% of pupils were taught by teachers who wished to have future CPD in integrating science subjects, pupil needs, pupil assessment, and critical thinking.

Pupil achievement

The level of deprivation has a much larger impact on pupil achievement than any other variable we considered. For the IDACI score, a unit change from lower levels of deprivation to higher levels of deprivation saw an average reduction in the score of -26.5 points (-71.2, 18.7). Home resources of the pupil, on the other hand, made a substantial and positive difference with pupils with a 1-unit higher level scored 17.0 points higher [16.8, 17.2]. This reflects and reinforces the advantage / disadvantage gap in young children and is more frequently reported in the later years of secondary education.

Prior achievement is a predictor of performance. For the score on the phonics test, a one-unit improvement led to an increased score on the science assessment of 3.2 points (2.6, 3.8).
The 'affect' is also important: pupils who enjoy learning science fare better on the TIMSS test and pupil confidence had a stronger positive association with a value of 8.6 points [8.3, 8.8] although of note is that a smaller group of pupils in England felt very confident compared with their international peers.

Approaches to teaching and learning science

More 'inquiry-based 'approaches were not always clearly associated with higher test scores and suggest much more complexity to appropriate teaching methods in primary science. When teachers ask pupils to *conduct experiments* or *investigations*, for those in the *some lessons* group was positively associated with test scores, but this was not significant [-2.0, 2.8], but when conducting experiments became more intensive in use, these groups had more negative associations with the test score. When teachers asked pupils to *interpret data* and / or *use evidence from experiments or investigations to support conclusions*, all the coefficients showed positive associations with the overall test scores, although not all were statistically significant.

Conclusions

The low participation rate and incomplete responses by teachers in England suggests we may not have a country-wide representative population. In turn, this limits the lessons learned or the understanding of primary science teaching and learning[2]. Given these caveats, from the analysis of the data we have been able to use, the following conclusions can be drawn:

- **Instructional time and clarity**

We have not found an association between instructional time or instructional clarity and achievement for pupils in England. In comparison, analysis of the international data shows that there is a 'sweet spot' of instructional time, so efforts to supplement science learning experiences at this level of schooling with additional out of school teaching time is not beneficial. The TIMSS report details that, 'internationally and within most countries, however, higher clarity was associated with higher average achievement' (Mullis, Martin et al., 2019).

- **Frequency of investigation, pupil enjoyment and interest in science**

We have found that specific aspects of classroom scientific investigative work are positively associated with achievement. When pupils reported greater frequency of teachers asking questions to promote thinking, stimulate discussions, assess their learning, this was positively associated with attainment. This augurs well for supporting teachers to develop a 'thinking classroom' and as recommended by the EEF Primary Science Guidance Report, particularly in 'cultivating reasoning and justification'. By contrast, emphasis on the procedural (planning,

[2] For a discussion of TIMSS sampling within England, please see Appendix B, p 239.

conducting) with frequent use of classroom experimental work was negatively associated with attainment.

- **Teachers**

TIMSS data showed that the highest achievement for Year Five pupils was associated with teachers in England who had a background of education in science.

Higher levels of attainment are also associated with teachers who said …pupils are better when they are:

- confident in learning science;
- asked to interpret data and use evidence;
- read textbooks and source materials;

We note that that the phonics score at the end of Year One is closely correlated with the Year Five TIMSS score, that KS2 data was unavailable and that the sampling science tests for KS2 were not undertaken in this year.

- **Professional development experiences: past and future**

Limited participation by teachers in England of the TIMSS teacher survey limits the lessons that can be learned from the analysis. In sum, we have not established a causal relationship between PD for elementary (primary) teachers and their pupils' attainment. The 'gap' in experience and future needs for CPD are not so much concerned with science content or curriculum but rather more general including the integration of science, the integration of technology, and meeting pupil needs.

Concluding remarks

However, the aspects that pupils bring to school themselves, the influence of their home background is much larger than any of the above factors. The IDACI score has a much larger effect on pupil achievement where this measure of deprivation is reflected in lower pupil achievement scores. As reported earlier this year (EPI, 2024), the attainment gap exists in early childhood and continued to widen throughout their schooling. This suggests that very early interventions to ameliorate the effect of economic disadvantage might result in more equitable attainment across pupils.

Introduction

This study uses data from the 2019 Trends in International Mathematics and Science Study (TIMSS), to better understand factors impacting pupil achievement in science in Grade Four (Year Five in England). TIMSS is an international assessment project conducted by the International Association for the Evaluation of Educational Achievement (IEA) and is focused on Mathematics and Science conducted at the fourth and eighth grade level (Year 5 and 9 within England). In 2019, 64 countries took part, and the TIMSS curriculum model compares the intended, implemented and attained curriculum with questionnaires targeted at the pupil, home, teacher and school principal level. To ensure comparisons across different countries and jurisdictions, TIMSS has a clearly defined approach to sampling from the school population agreed between each national coordinator and TIMSS sampling experts.

The two-stage sampling process of selecting representative schools and classes ensures a TIMSS precision target of pupil achievement resulting in a sample of approximately 4000 pupils in 150 schools. Most countries select one class per school in the sample, to ensure a proportional representation of pupils in the selected schools. Science assessment in TIMSS has been constructed on frameworks developed by the participating countries for each curriculum area and for each grade. The cognitive items aimed to assess pupils' achievement in the domains of knowing, reasoning and application (Martin et al., 2020). These are measured using a science test and estimated using a measurement model producing five plausible values, representing the range of pupil performance. A comparative secondary data analysis of pupils in primary schools provides an international perspective to identify the relative strengths of factors within pupils, schools and classrooms and determine associations between pupil achievement and interest in science. In this report, we also explore how teachers' professional development, years of teaching, and instructional practices are associated with achievement in science.

The research questions that guided this study are:

RQ1. What are the professional development experiences of primary teachers in England in teaching science. In particular,

I. Is there a relationship between the time spent, and types of CPD on science specific CPD and pupil achievement?
II. What types of CPD do primary teachers identify for their future needs to support them teaching science?

RQ2. What is the association of instructional time of science in primary schools and pupils' achievement? What is the association of instructional clarity and pupils' achievement?

RQ3. What is the association between frequency of teachers' use of science investigation, pupils' investigations and pupil enjoyment of, interest in and achievement?

Teachers were asked in the teacher survey about the school's emphasis on academic success, school environment; pupils' behaviour, rules, safety, fairness; about being a teacher, including how they felt (meaning and purpose, enthusiasm) workload (teaching hours, pressure from parents, administrative tasks, preparation for class and curriculum changes).

To understand how teachers work in the Year Five class, teachers were asked about how often they were able to relate lessons to pupils' daily life, ask pupils to explain their answers, ask pupils to decide their own problem-solving procedures, encourage classroom discussions among pupils and link new content to pupils' prior knowledge. Teachers were asked about the limiting factors on their teaching such as pupil absence, disruption, pupils' lack of sleep/ nutrition, lack of interest and language and behavioural difficulties. There were questions specifically about the teaching and learning of mathematics and science in the classroom. Finally, there were questions that asked about access to and participation in CPD and identification of CPD needs with a focus on content, pedagogy, curriculum integrating technology, improving critical thinking skills, assessment, addressing individual needs and integration with other subjects.

Primary Science in England

The home-grown STEM pipeline needs support from the early years to ensure future economic prosperity, improve scientific literacy and citizenship with implications for teaching and learning of science. Despite science being a mandated core subject in primary schools, government school inspection reports that primary science has been ‘downgraded’ since whole cohort assessment of science at key stages ceased (Office for Standards in Education, 2019), and pupils experienced varying and varied levels of science learning (CFE, 2019: Ofsted, 2023). It may be that primary teachers in England feel less well-prepared to provide a rich science learning experience for pupils, in terms of their own scientific knowledge, support for and confidence in teaching science as well as time constraints within the school curriculum. If this is the case, this may translate into a curriculum narrow in terms of content and coherence raising questions about time spent, aspects of the science curriculum covered, and the experiences pupils have in learning science, particularly practical and investigative science. Pre-service and in-service PD School leaders and head teachers may lack provision of ‘strategic professional development related to science subject pedagogy’ (Bianchi, Whittaker, & Poole, 2021). A Wellcome-initiated survey reported teacher anxiety about teaching some topics, including forces, light, electricity and evolution (Stubberfield, 2021). In schools with strong science leadership, Ofsted and Wellcome reported coherence of the primary curriculum, purpose of the curriculum and learning activities. Given the breadth and scope of subjects that primary teachers cover in their classes, which may alter year by years, targeted professional development might be difficult to access.

Teacher qualification and professional development in science

Professional development (PD) of teachers supports the development of skills and knowledge in ensuring quality and standards of teaching. It includes both in-service and pre-service teacher education. Common consensus assumes that professional development of teachers builds capacity through greater subject knowledge, teaching skills, instructional practices, and translating into improved attainment and achievement of pupils. Effects of high-quality teaching on pupil learning and achievement are well documented (Darling-Hammond, 2010; Desimone, 2009; Hattie, 2003; Timperley, 2008). Recognising that effective PD needs to be well-designed, targeted and implemented (Darling-Hammond, Hyler & Gardner, 2017; de Vries, Dimosthenous, Schildkamp & Visscher, 2022), a ‘one size fits all’ nature of teacher PD is questionable. The consensus view about collaborative, sustained (rather than repeated) is widely seen as providing ‘best practice’ for PD providers and school leaders although the most recent evidence suggests that the evidence for this is contested, even flawed (Sims & Fletcher- Wood, 2021). ‘Instructional coaching’, for example, is positively associated with pupil outcomes (achievement), providing evidence that the mechanism of teacher changing their practice is associated with beneficial outcomes for pupils. On the other hand, the evidence for ‘collaboration’ is lacking. If and when the focus of PD is on pupil outcomes, that researchers and policy makers will make use of both ‘evidence of mechanism and evaluations of specific PD interventions which include those mechanisms’ (Sims & Fletcher- Wood, 2021, p. 57-58).

PD programmes involving elementary teachers of science report positive findings on teachers’ subject knowledge, confidence and practices (Maeng, Whitwirth, Bell & Sterling, 2020) and an intervention to improve teachers’ content knowledge associated with pupil achievement on high-stakes tests (Diamond, Maerten-Rivera, Rohrer, & Lee, 2014). A three- year PD intervention in Indiana, USA, reported that both recency of, and continuous engagement with science-specific PD was associated with pupil attainment (Mutch-Jones, Hicks, & Sorge, 2022). An approach combining analysis-of-practice and science content deepening with a PD programme is reported to be effective at improving elementary pupils’ science content knowledge (Taylor, Roth, Wilson, Stuhlsatz, & Tipton, 2017) with caveats about the fidelity and generalisability of the PD intervention.

The Trends in International Mathematics and Science Survey 2019 collected data from participating teachers, schools, families and pupils. These included context questions and teachers were asked for information about their academic and professional backgrounds (level of education, their main focus of study during tertiary education, how much science was in their pre-service programme of study, for example), classroom resources, teaching strategies (instructional approaches) and attitudes towards teaching. We have used responses to these questions in addition to ordinal questions addressing confidence in teaching science, lesson delivery, science topics taught, how well-prepared they are for teaching aspects of science, and a series of dichotomous questions on science professional development. These are important to explore in depth given the anecdotal and disappointing reports from Ofsted about coherence and coverage of the science curriculum in primary schools. TIMSS data showed that the highest achievement for Year Five pupils was associated with teachers in England who had a background of education in science (Richardson et al., 2020). One assumption is that background knowledge in science enables teachers to plan for a cohesive and deeper learning experience for their pupils. This study reports on the frequency and types of science specific CPD, how much they have been able to access in the past two years and what CPD needs they have to deliver a balanced science curriculum.

This analysis uses The International Mathematics and Science Survey 2019 to explore teachers' professional development in England. The TIMSS 2019 guide states that:

"The teachers in the TIMSS 2019 International Database do not constitute representative samples of teachers in the participating countries. Rather, they are the teachers of nationally representative samples of pupils. Therefore, analyses with teacher data should be made with pupils as the units of analysis and reported in terms of pupils who are taught by teachers with a particular attribute".

Following this guidance, this analysis expresses the proportion of children taught science by teachers who had or hope to have professional development in a variety of topics.In considering the different experiences teachers have in accessing professional development, TIMSS data show how much this varies across countries. The particular subset focusing on England will be considered later. Figure 1 shows how teachers in participating countries report the hours of CPD undertaken in the previous two years.

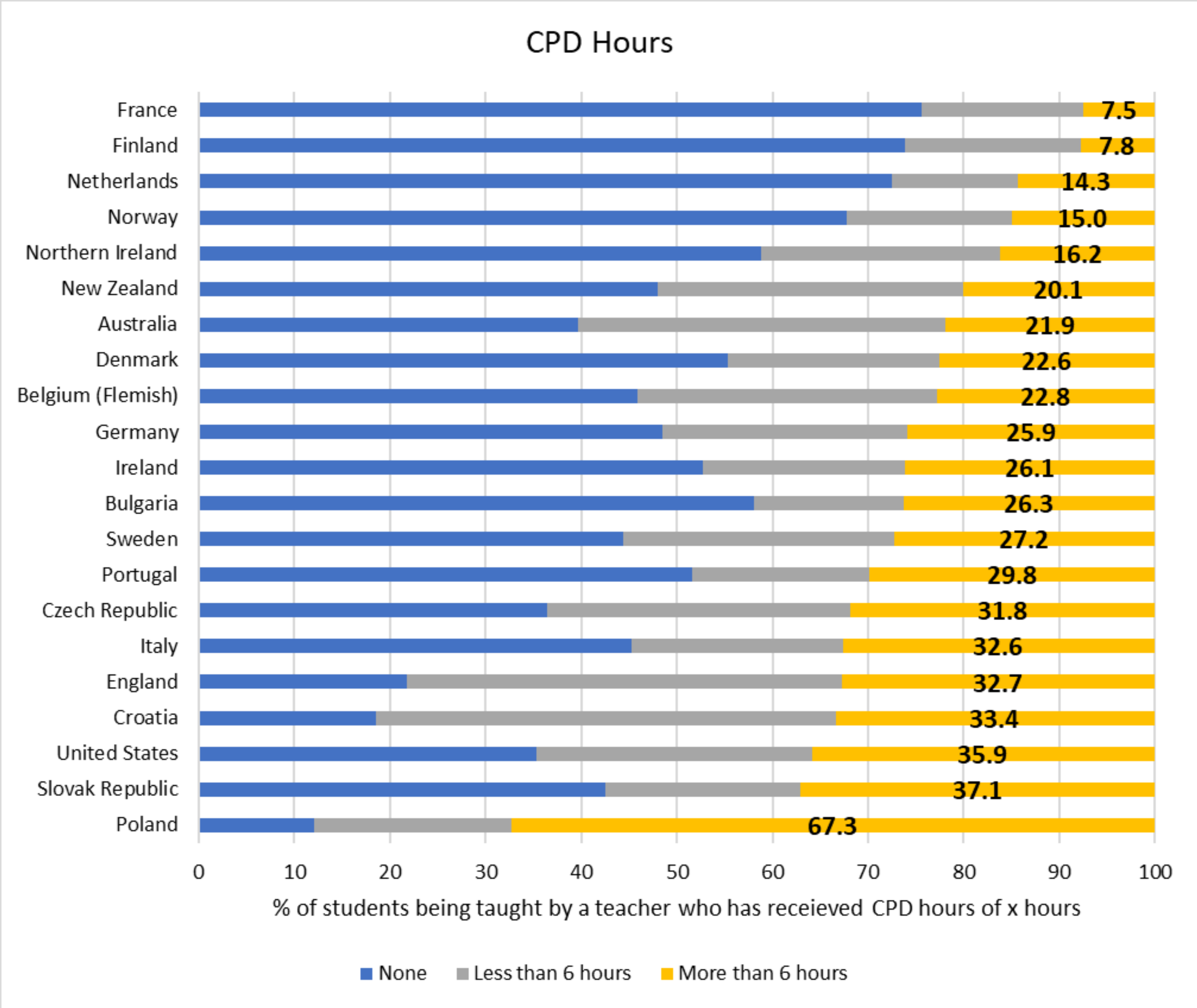


Figure 1. Professional Development Hours by Country (in the preceding two years)

Figure 1 shows the proportion of pupils in each country who are taught by teachers who have received no hours of professional development training, up to 6 hours of CPD and more than 6 hours of CPD. Teachers in Poland are more likely than teachers in other countries to have received more than 6 hours of CPD. In France, Finland and the Netherlands more than 70% of pupils were taught by teachers who had received no professional development training.

Teachers were asked about past and anticipated (or identified) professional development. In England, even though the teacher survey data is incomplete, we can see that more than half of the pupils taught had teachers reporting participation on CPD with a focus on science curriculum. In terms of future CPD needs, teachers identified aspects where this is needed. Figure 2 shows the past (two years) experienced CPD and future CPD needs of teachers in England.

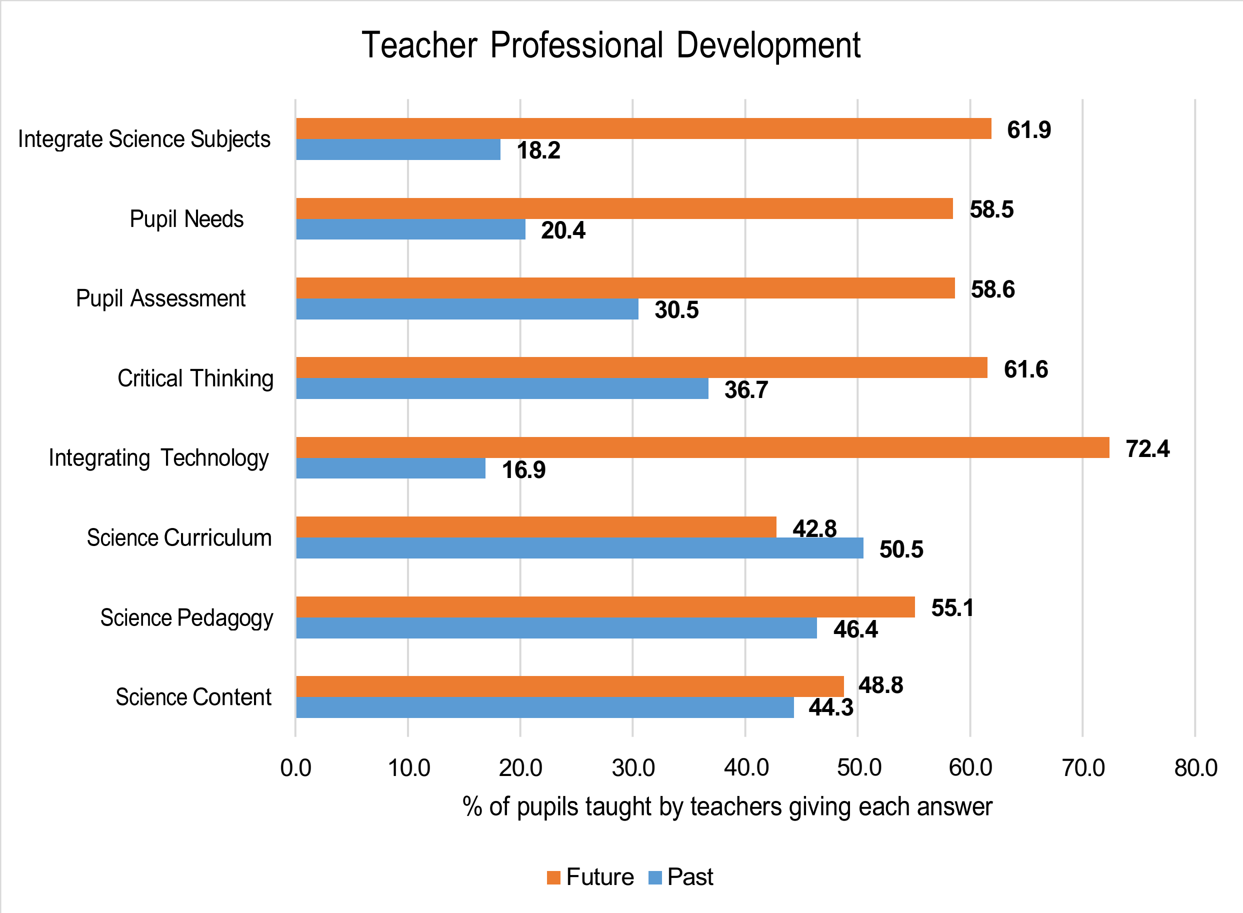


Figure 2. Teachers' CPD experiences and their future needs (England)

These data show that teachers in England express a greater wish for continuing professional development in some areas, compared to others. Just over half the teachers completing the survey have taken part in CPD with a focus of the science curriculum. 72.4% of pupils were taught by a teacher who wishes to have future professional development opportunities in integrating technology, and only 16.9% of pupils were taught by a teacher who had already had training in this area. Around 60% of pupils were taught by teachers who wished to have future CPD in integrating science subjects, pupil needs, pupil assessment, and critical thinking. In these areas fewer than 2 in 5 pupils were taught by teachers who had had training in these areas in the past. The 'gap' in experience and future needs for CPD are not so much concerned with science content or curriculum but rather more general including the integration of science, the integration of technology, and meeting pupil needs such as?.

How this compares with teachers in other countries is of interest. Just looking at one aspect of this, participation of CPD in the past two years where science content was the focus shows considerable variation.

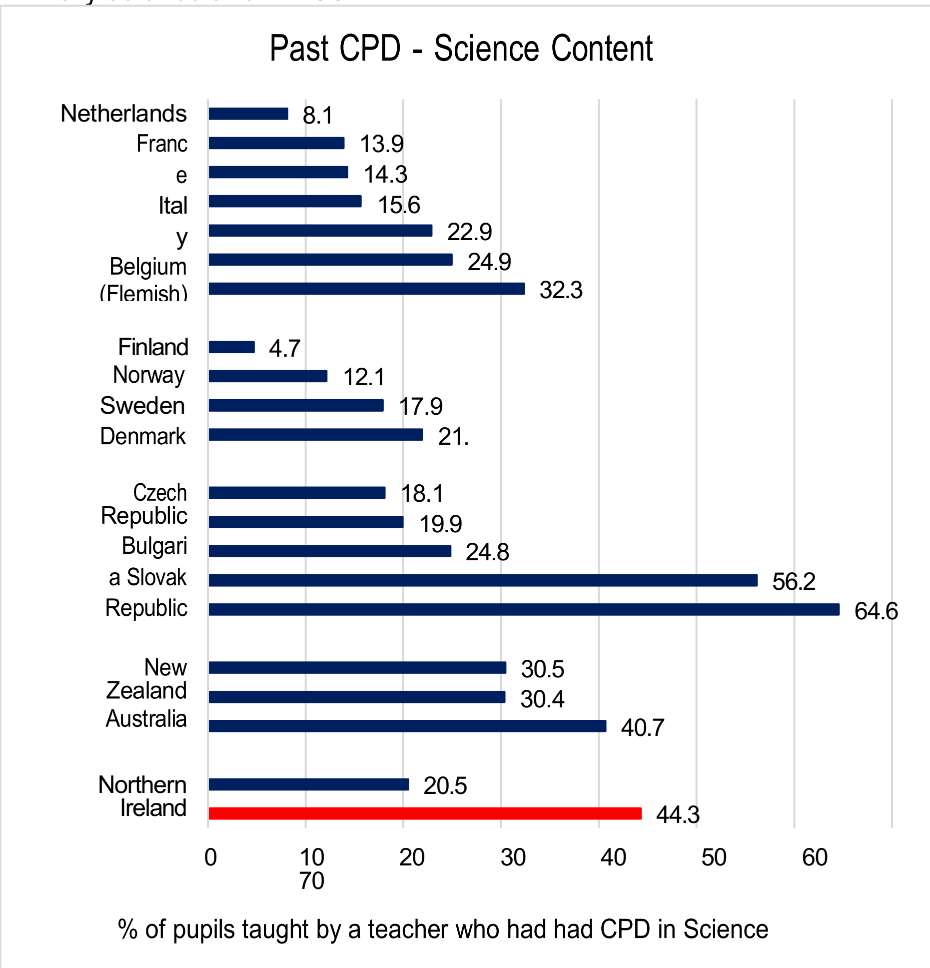


Figure 3. Teachers' CPD experiences with focus on science content

Nearly 44% of pupils have primary teachers in England who have experienced CPD in science. This finding resonates with the analysis of the Wellcome Report (2019) where 52% of science leaders had experienced external CPD in science and 5% of teachers had accessed science-specific CPD.

Table 1. The demographic profile of Year Five teachers in England

	% of pupils taught by teachers
Female	43,
1-5 years' teaching	23
6-10 years' teaching	15
11-20 years' teaching	18
21+ years' teaching	7

Table 2. Teacher Age (total 3396)

Teacher age	n	% of pupils taught by teachers
Under 25	139	4
25-29	671	20
30-39	695	20
40-49	435	13
50-59	82	2
60+	70	2

There is a considerable amount of missing data on teacher's demographic information. Between

37% and 38% of pupils have no information on the gender, age or teaching experience of their teachers, and 43% of pupils are reported to be taught by women and 19% taught by men but we do not have any information for the ‘missing data’ where pupils were taught by teachers who did not complete this part of the questionnaire.

Roughly 1 in 5 pupils are taught by teachers aged 25 to 29 and another 20% taught by teachers who are 30 to 39 years old. 22% of pupils in the survey are taught by teachers who have 5 years teaching experience or less. 15% are taught by teachers with 6 to 10 years’ experience and 18% by someone with 11 to 20 years of experience working as a teacher. Just under 7% were taught by teachers with 21 years or more experience in the job. By and large, the profile of a primary Year Five teacher is likely to be represented by a female and young.

Of greatest concern is the robustness of the dataset for teachers in England compared with other countries. Indeed, disappointingly England’s returns saw an average missingness of 52% across the variables in the international data model below. Furthermore, there were a couple of columns of 100% missing data, which resulted in England being dropped from the analysis despite pupils completing their tests and questionnaires.

with pupil achievement in England: pupils and teachers

We wanted to explore the various factors associated with achievement on TIMSS, using pupil demographic data, their self-reported responses to questions about their science learning in school and their teachers’ responses about their teaching and CPD experiences (Richardson et al., 2020)

In England, 3,396 primary pupils from 139 primary schools participated, including eight independent schools and most being ‘community’ or state-funded schools. An in-depth examination was conducted with an extract of the TIMSS data for England linked to the National Pupil Database (NPD), held by the by the International Statistics division within the Department for Education (DfE). This allows for analysis of factors including free school meals (FSM), ethnicity as well as allow for a quasi-longitudinal study on pupil progress from the Foundation Stage Profile. The DfE were able to link 3319 TIMSS participating primary school pupils to their existing records which provided important measures of deprivation for the pupil (everFSM) whether pupils were ever recipients of the free school meal entitlement), and the Income Deprivation Affecting Children Index (IDACI) which is a postcode level measure of the proportion of children under the age of 16 that live in low income homes. Measures of prior achievement were more limited given the age of the pupils at the time of the 2019 TIMSS fieldwork, with follow on tests at Key Stage 2 (KS2) cancelled due to Covid. We used the recorded Phonics mark from the screening check in Year 1 to gain some sense of the level of their prior achievement.

The model (includes international data)

Our modelling proceeded on the basis of incremental improvements working from a basic null (or empty) model with no predictors, to a varying intercept model where indexes were added to account for clustering at the class, school and country level. We then increased the complexity of the model adding further pupil, teacher and school level variables and assessing model fit. Our outcome of interest was the TIMSS Grade Four (England, Year 5) score which is centred on a score of 500 prior to including additional explanatory predictors.

The final model is built as follows.

- Firstly, we included predictors to account for pupil background which comprised of a binary sex of pupil variable (where the base case was female) and a continuous home resources for learning scale. In particular, the home resources scale asked pupils and parents on the number of books in the home, home study supports, the highest education level of either parent, and the highest occupational level of either parent.
- Secondly, we included three measures of pupil perceptions on whether they like learning science, their confidence in science and the instructional clarity in science lessons. In regards to instructional clarity, pupils were asked about aspects of their teachers’ instruction: ‘whether they knew what their teacher expects them to do, whether their

teacher is easy to understand, has clear answers to their questions, is good at explaining things, and does a variety of things to help them learn … and explains a topic again when pupil do not understand' (TIMSS, p.475).

- We also controlled for the amount of instructional time received by the pupils which was mean-centred, and whether they received any additional instruction in the form of further science lessons (such as with individual tutoring)
- Thirdly, at the teacher level we included 10 variables examining how science lessons are taught which had the following categories - "Every or almost every lesson" (base category), "About half the lessons", "Some lessons" and "Never". These variables examined pupils listening to the teacher explain; observe phenomena, the being demonstrated an experiment, planning experiments, conducting experiments, being presented data, interpreting data, using evidence, reading textbooks, memorising facts and doing fieldwork
- Fourthly, we included a series of discrete variables (with three categories – a lot, some and none which was set as the base case) that measured the importance that teachers attach to specific assessment strategies. These were observing pupils as they work; asking pupils to answer questions during class; short, regular written assessments; longer tests; and long-term projects.
- Lastly, we also included a series of dichotomous variables to understand the impact of various professional development courses taken prior to the TIMSS survey. These included training on science content, science pedagogy, science curriculum, integrating technology, standard critical thinking, science assessment, past pupil needs, and integrating science subjects into the curriculum.

The equation for the model:

$$\beta_{0ijkl} + \beta_1 SexMale_{ijkl} + \beta_2 HomeResources_{ijkl} + \beta_3 InstructionalClarity_{ijkl}$$
$$+ \beta_4 Studentslikelearningscience_{ijkl} + \beta_5 StudentConfidenceinScience_{ijkl}$$
$$+ \beta_6 Receivesextralessonsinscience_{ijkl} + \beta_7 Timespentoninstruction_{ijkl}$$
$$+ \beta_8 Studentslistentomeexplain2_{ijkl} + \beta_9 Studentslistentomeexplain3_{ijkl}$$
$$+ \beta_{10} Studentslistentomeexplain4_{ijkl} + \beta_{11} Observephenoma2_{ijkl}$$
$$+ \beta_{12} Observephenoma3_{ijkl} + \beta_{13} Observephenomena4_{ijkl}$$
$$+ \beta_{14} Demonstrateexperiment2_{ijkl} + \beta_{15} Demonstrateexperiment3_{ijkl}$$
$$+ \beta_{16} Demonstrateexperiment4_{ijkl} + \beta_{17} Planexperiment2_{ijkl}$$
$$+ \beta_{18} Planexperiment3_{ijkl} + \beta_{19} Planexperiment4_{ijkl}$$
$$+ \beta_{20} Conductexperiment2_{ijkl} + \beta_{21} Conductexperiment3_{ijkl}$$
$$+ \beta_{22} Conductexperiment4_{ijkl} + \beta_{23} Presentdata2_{ijkl} + \beta_{24} Presentdata3_{ijkl}$$
$$+ \beta_{25} Presentdata4_{ijkl} + \beta_{26} Interpretdata2_{ijkl} + \beta_{27} Interpretdata3_{ijkl}$$
$$+ \beta_{28} Interpretdata4_{ijkl} + \beta_{29} Useevidence2_{ijkl} + \beta_{30} Useevidence3_{ijkl}$$
$$+ \beta_{31} Useevidence4_{ijkl} + \beta_{32} Readtextbooks2_{ijkl} + \beta_{33} Readtextbooks3_{ijkl}$$
$$+ \beta_{34} Readtextbooks4_{ijkl} + \beta_{35} Memorisefacts2_{ijkl} + \beta_{36} Memorisefacts3_{ijkl}$$
$$+ \beta_{37} Memorisefacts4_{ijkl} + \beta_{38} Dofieldwork2_{ijkl} + \beta_{39} Dofieldwork3_{ijkl}$$
$$+ \beta_{40} Dofieldwork4_{ijkl} + \beta_{41} Specialisationinscience_{ijkl}$$
$$+ \beta_{42} Importance{:}\ Obsstudents2_{ijkl} + \beta_{43} Importance{:}\ Obsstu3_{ijkl}$$
$$+ \beta_{44} Importance{:}\ Obsstu4_{ijkl} + \beta_{45} Importance{:}\ askingstu2_{ijkl}$$
$$+ \beta_{46} Importance{:}\ askingstu3_{ijkl} + \beta_{47} Importance{:}\ askingstu4_{ijkl}$$
$$+ \beta_{48} Importance{:}\ shortassess2_{ijkl} + \beta_{49} Importance{:}\ shortassess3_{ijkl}$$
$$+ \beta_{50} Importance{:}\ shortassess4_{ijkl} + \beta_{51} Importance{:}\ longertests2_{ijkl}$$
$$+ \beta_{52} Importance{:}\ longertests3_{ijkl} + \beta_{53} Importance{:}\ longertests4_{ijkl}$$
$$+ \beta_{54} Importance{:}\ longtermproj2_{ijkl} + \beta_{55} Importance{:}\ longtermproj3_{ijkl}$$
$$+ \beta_{56} Importance{:}\ longtermproj4_{ijkl} + \beta_{57} PDPast{:}\ ScienceContent_{ijkl}$$
$$+ \beta_{58} PDPast{:}\ SciencePedagogy_{ijkl} + \beta_{59} PDPast{:}\ Curriculum_{ijkl}$$
$$+ \beta_{60} PDPast{:}\ Technology_{ijkl} + \beta_{61} PDPast{:}\ criticalthinking_{ijkl}$$
$$+ \beta_{62} PDPast{:}\ assessment_{ijkl} + \beta_{63} PDPast{:}\ studentneeds_{ijkl}$$
$$+ \beta_{64} PDPast{:}\ integratesciencesubjects_{ijkl} + u_{0j} + v_{0k} + f_{0l}$$

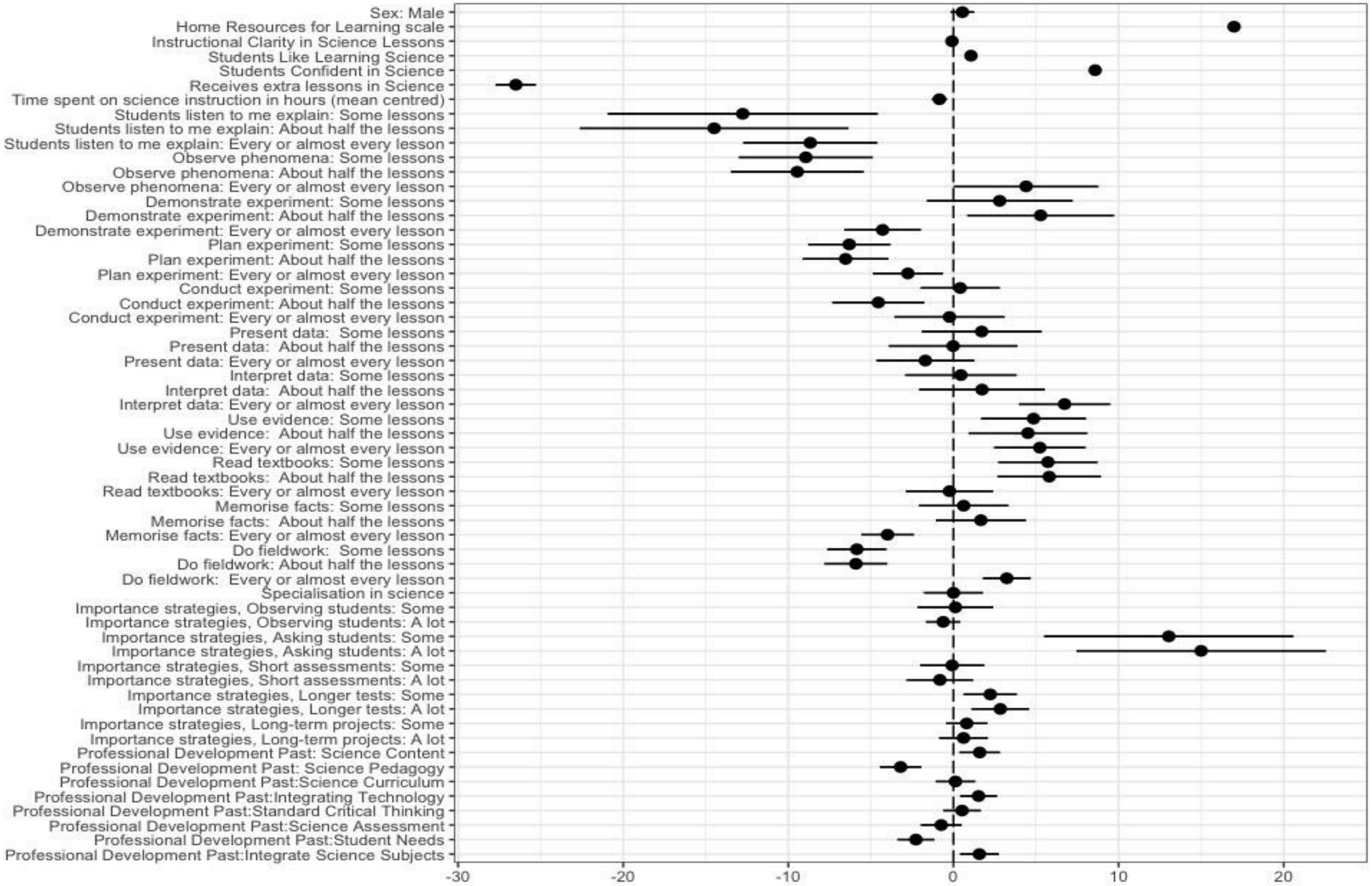
Missing data

'Missingness' within the international TIMSS is a complex issue as not all participating countries completed all of the surveys. This is a large data set, built from individual pupils, nested within classes, nested within schools, nested within countries so we have relied on complete case analysis. Unfortunately, sophisticated approaches such as Full Information Maximum Likelihood or multiple imputation are generally limited to two-level designs (such as pupils nested in schools).

How home resources and teachers asking pupil questions supports pupil learning

Figure 4 shows how different factors are associated with pupil attainment on the TIMSS Grade Four assessment. The intercept showed that the average pupil in the average class, school and country scored 499.0 points [481.9, 516.0]. For the demographic predictors, male pupils scored marginally higher with a coefficient of 0.5 [-0.2, 1.2], showing little in the way of a gender effect. Home resources of the pupil on the other hand made a substantial difference with pupils with a 1-unit higher level scored 17.0 points higher [16.8, 17.2]. This reflects and reinforces the advantage / disadvantage gap in young children and is more frequently reported in the later years of secondary education. The Education Policy Institute (2023, 2024) has reported that the disadvantage gap has grown since 2019 and continues growing throughout schooling from early years with a gap of 4.8 months, and at the end of KS2, of 10.3 months and KS4, the gap of 18.8 months.

Figure 4.


Sex: Male
Home Resources for Learning scale
Instructional Clarity in Science Lessons
Students Like Learning Science
Students Confident in Science
Receives extra lessons in Science
Time spent on science instruction in hours (mean centred)
Students listen to me explain: Some lessons
Students listen to me explain: About half the lessons
Students listen to me explain: Every or almost every lesson
Observe phenomena: Some lessons
Observe phenomena: About half the lessons
Observe phenomena: Every or almost every lesson
Demonstrate experiment: Some lessons
Demonstrate experiment: About half the lessons
Demonstrate experiment: Every or almost every lesson
Plan experiment: Some lessons
Plan experiment: About half the lessons
Plan experiment: Every or almost every lesson
Conduct experiment: Some lessons
Conduct experiment: About half the lessons
Conduct experiment: Every or almost every lesson
Present data: Some lessons
Present data: About half the lessons
Present data: Every or almost every lesson
Interpret data: Some lessons
Interpret data: About half the lessons
Interpret data: Every or almost every lesson
Use evidence: Some lessons
Use evidence: About half the lessons
Use evidence: Every or almost every lesson
Read textbooks: Some lessons
Read textbooks: About half the lessons
Read textbooks: Every or almost every lesson
Memorise facts: Some lessons
Memorise facts: About half the lessons
Memorise facts: Every or almost every lesson
Do fieldwork: Some lessons
Do fieldwork: About half the lessons
Do fieldwork: Every or almost every lesson
Specialisation in science
Importance strategies, Observing students: Some
Importance strategies, Observing students: A lot
Importance strategies, Asking students: Some
Importance strategies, Asking students: A lot
Importance strategies, Short assessments: Some
Importance strategies, Short assessments: A lot
Importance strategies, Longer tests: Some
Importance strategies, Longer tests: A lot
Importance strategies, Long-term projects: Some
Importance strategies, Long-term projects: A lot
Professional Development Past: Science Content
Professional Development Past: Science Pedagogy
Professional Development Past:Science Curriculum
Professional Development Past:Integrating Technology
Professional Development Past:Standard Critical Thinking
Professional Development Past:Science Assessment
Professional Development Past:Student Needs
Professional Development Past:Integrate Science Subjects
-30
-20
-10
0
10
20

For the pupil perceptions and experiences of science instruction and learning, the impact was mixed. For *instructional clarity*, this variable aimed to capture pupil perceptions about their class teacher: whether pupils know what the teacher expects them to do, is easy to understand, has clear answers to pupils' questions, is good at explaining things, does a variety of things to help them learn and explains a topic again when they don't understand (MS8). There was a very small negative association for *instructional clarity* with a coefficient of -0.1 [-0.3, 0.1].

For the pupils who '*like learning science*', there was a small positive association of 1.1 points [0.8, 1.3]. This is an aggregate variable (MS7), capturing whether pupils enjoy learning science, like doing science experiments and look forward to learning science, for example. Pupils who enjoy learning science fare better on the TIMSS test. *Pupil confidence* was an aggregate variable (MS9) capturing aspects (I usually do well in science, science is harder for me than many of my classmates, I am just not good at science, my teacher tells me I am good at science, science is harder for me than any other subject and science makes me confused), some of which were reverse scored. TIMSS identified three categories of confidence. Our analysis shows that confidence had a stronger positive association with a value of 8.6 points [8.3, 8.8] and a smaller group of pupils in England felt very confident compared with their international peers.

The international data set show variation in *instructional time* in science throughout a school year. The data submitted for England were insufficient to use as there was a threshold of at least 40% of data for pupils to be analysed. Receiving additional lessons in science was strongly (and surprisingly) negatively associated with a coefficient of -26.5 [-27.7, -25.3], and time spent on science instruction was also negatively associated with a value of -0.9 [-1.3, - 0.4]. This appears to be counterintuitive, but the international analysis shows no clear relationship between hours of formal instruction and pupil achievement. It may be that pupils who have lower levels of achievement are offered additional lessons outside of the normal school day.

We used responses from the teacher questionnaire to explore how different teaching strategies were associated with pupil scores. A passive method of instruction, when pupils listen to the teacher, resulted in lower test scores on average. What is of note here, is this used the teacher report matched with the pupil score in that class. Pupils who listened to the teacher explain in *some* lessons scored 12.8 lower [-20.9, -4.6] on average than the base case. [The *never* category is definitely smaller across selection of variables, but not that rare. It does not change the slope of the model but may help estimate the intercept with more precision. However, choosing a base case at either end can help with the interpretation. The average number of cases across the relevant variables in the base category is approximately 7888 with a standard deviation of 6836, minimum of 546 and maximum of 24168.]

Pupils who listened to teachers explain in *half the lessons* scored 14.5 points lower [-22.6, -6.4] on average than the base case. Lastly pupils who listened to the teacher explain in *most or all lessons* scored 8.7 points lower [-12.7, -4.6] on average than the base case. The spread of data across scores shows that it is not easy to identify a causal relationship between teachers explaining and pupil achievement. That it varies so much warrants greater exploration. It may be that the limited question (*how often do you ask your pupils to listen to me explain new science content*) is not capturing qualitatively teachers' use of explanations about science or as part of a broader pedagogical approach.

More 'inquiry-based 'approaches were not always clearly associated with higher test scores and suggest much more complexity to appropriate teaching methods in primary science. *Observing phenomena* in *some* lessons and about *half* the lessons were negatively associated with the test score. When teachers asked pupils to *observe phenomena* in *some* lessons, pupils scored -8.9 points lower [-13.0, -4.9] on average than the base case. For those *observing phenomena* in about *half* the lessons scored 9.5 points lower [-13.5, -5.4] on average than the base case. However, when pupils were asked to *observe phenomena* in *every lesson*, they scored on average 4.4 points higher [0.0, 8.8] than the base case. This suggests that teachers asking pupils to notice, observe natural phenomena and describe what they see can be beneficial to pupils. We don't know whether it is a group activity, individual or whole class experience but this lends itself to further exploration.

When teachers ask pupils to *watch the teacher demonstrate an experiment or investigation*, in *some* or about *half* the lessons, this was positively associated with test scores. For the former, pupils scored on average 2.8 points higher [- 1.6, 7.2] than the base case, and for the latter, pupils scored on average 5.3 points higher [0.8, 9.7] than the base case. However, those teachers demonstrating experiments in most or all lessons saw their pupils on average score 4.3 points lower [-6.6, -2.0] than the base case. There is a slight difference in teachers asking pupil to notice and describe rather than watch a teacher demonstrate an activity which may explain the different outcomes. Additionally, we do not know the criteria that teachers were asked to use in distinguishing between observing phenomena and watching the teachers demonstrate an experiment, other than the survey questionnaire, so this aspect of teacher and teaching strategy needs further unpacking.

When teachers ask pupils to design or plan experiments or investigations, pupil responses in all groups (in some lessons, about half the lessons, and every or almost every lesson) were all negatively associated with test scores. For planning experiments in some lessons, pupils scored an average of 6.3 points lower [-8.8, -3.8] than the base case. For those in the about half the lessons group this was 6.5 points lower [-9.1, -3.9], and for those in every or almost every lesson group this was 2.8 points lower [-4.9, -0.6] on average and a small range of data points. This might reflect the difficulty of the task: we might expect a good plan to reflect experience such as familiarity with experimental protocol, specific subject knowledge and scaffolding from teachers.

When teachers ask pupils to conduct experiments or investigations, for those in the some lessons group was positively associated with test scores, but this was not significant [-2.0, 2.8], but when conducting experiments became more intensive in use, these groups had more negative associations with the test score. For those where conducting experiments formed about half the lessons, they averaged 4.6 points [-7.3-1.8] lower than the base case. For those where conducting experiments were in every or almost every lesson, they scored 0.2 points lower on average, although this was not significant with an interval of -3.6 to 3.1. Again, this is likely to reflect experience of investigative approaches in science.

When teachers asked pupils to present data followed a similar pattern to the previous variable (conducting an investigation / experiment) with those in the group. Where pupils were presented data in some lessons, they scored 1.7 points higher [-1.9, 5.3] on average than the base case. However, those where presenting data took place in about half the lessons scored little different to the base case [-3.9, 3.9], and those where presenting data was in every or almost every lesson on average scored 1.7 points lower [-4.7, 1.3] than the base case, although again it was not significant. Although these findings appear to be counter to current thinking about teaching and learning science, we note that this pattern is also observed in other studies, for example, using PISA data (Jerrim, Oliver & Sims, 2023; Oliver, McConney, Woods-McConney, 20 21). It might be that if the focus on the procedural rather than the cognitive aspects of investigative work, pupils also focus on the 'doing' instead of connecting to theory, abstracting and developing scientific explanations and 'making sense' of the experimental work.

When teachers asked pupils to interpret data and / or use evidence from experiments or investigations to support conclusions, all the coefficients showed positive associations with the overall test scores, although not all were statistically significant. For the group where pupils interpreted data in every or almost every lesson, they scored 6.7 points higher [4.0, 9.5] on average than the base case. For the group who used evidence in some lessons scored an average of 4.9 points higher [1.7, 8.0] than the base case. This was similar for the group where using evidence took place in about half the lessons with an average of 4.5 points higher [0.9, 8.1] than the base case, and marginally higher with the group using evidence in every or almost every lesson with an average of 5.2 points higher [2.5, 8.0] than the base case.

When teachers asked pupils to read textbooks or source materials in some and about half the lessons, this was positively associated with test scores with pupils scoring on 5.7 [2.7, 8.7] and

5.8 [2.7, 8.9] points higher than the base case on average. Asking pupils to memorise facts was mostly non-significant, although those groups that memorised facts in every or almost every lesson scored 4.0 points lower [-5.6, - 2.4] on average than the base case. Finally, for carrying out fieldwork outside the class, only the group which did this in every or almost every lesson was significant, and they averaged 3.2 points higher [1.8, 4.7] than the base case. It seems surprising that this strategy features frequently (in every or almost every lesson) in primary science lessons.

Variable	Probability of Observation
Science Score Plausible Value 1	1.00
Science Score Plausible Value 2	1.00
Science Score Plausible Value 3	1.00
Science Score Plausible Value 4	1.00
Science Score Plausible Value 5	1.00
Female	0.97
EverFSM	0.97
IDACI Score	0.97
Phonics mark	0.93
Instructional clarity in science	0.91
Pupils like learning science	0.92
Pupils confident in science	0.92
Pupils listen to me explain	0.48
Pupils observe phenomena	0.47
Pupils are demonstrated experiment	0.46
Pupils plan experiments	0.48
Pupils conduct experiments	0.48
Pupils are presented data	0.47
Pupils interpret data	0.46
Pupils use evidence	0.45
Pupils read textbooks	0.46
Pupils memorise facts	0.46
Pupils do field work	0.47
Professional development course attended on science content	0.27
Professional development course attended on science pedagogy	0.27
Professional development course attended on science curriculum	0.26
Professional development course attended on integrating technology	0.27
Professional development course attended on standard critical thinking	0.28
Professional development course attended on science assessment	0.28
Professional development course attended on pupil needs	0.28
Professional development course attended on science subjects	0.28

Table 3. Probability of observation

For teachers' assessment strategies (S6), teachers were asked on a three-point scale (a lot, some or none) about observing pupils as they work, asking pupils to answer questions during class, short regular written assessments, longer tests and long-term projects. When teachers asked pupils questions to answer during class, those who in the some or a lot groups saw the largest increases in test scores with some seeing 13 points higher [5.5, 20.6] score on average than the base case and a lot group saw test scores 15 points higher [7.5, 22.5] on average than the base case. Longer tests also showed statistically significant coefficients for using some longer tests (2.23 [0.6, 3.8]) and a lot (2.8 [1.1, 4.6]).

Lastly, for the past professional development training that teachers have undertaken, these all produced very small differences in TIMSS scores. PD with the focus on science content had a coefficient of 1.6 point [0.4, 2.8]; science pedagogy was associated with lower test scores with an average of -3.2 points [-4.5,-1.9]; integrating technology was associated with a marginally higher test score with average of 1.5 points [0.4, 2.6]; pupil needs was associated with lower test scores, with a coefficient of -2.3 [-3.4, -1.2]; integrating science subjects was associated with higher test scores, with a coefficient of 1.6 [0.4, 2.7].

Applying the model to data from England only

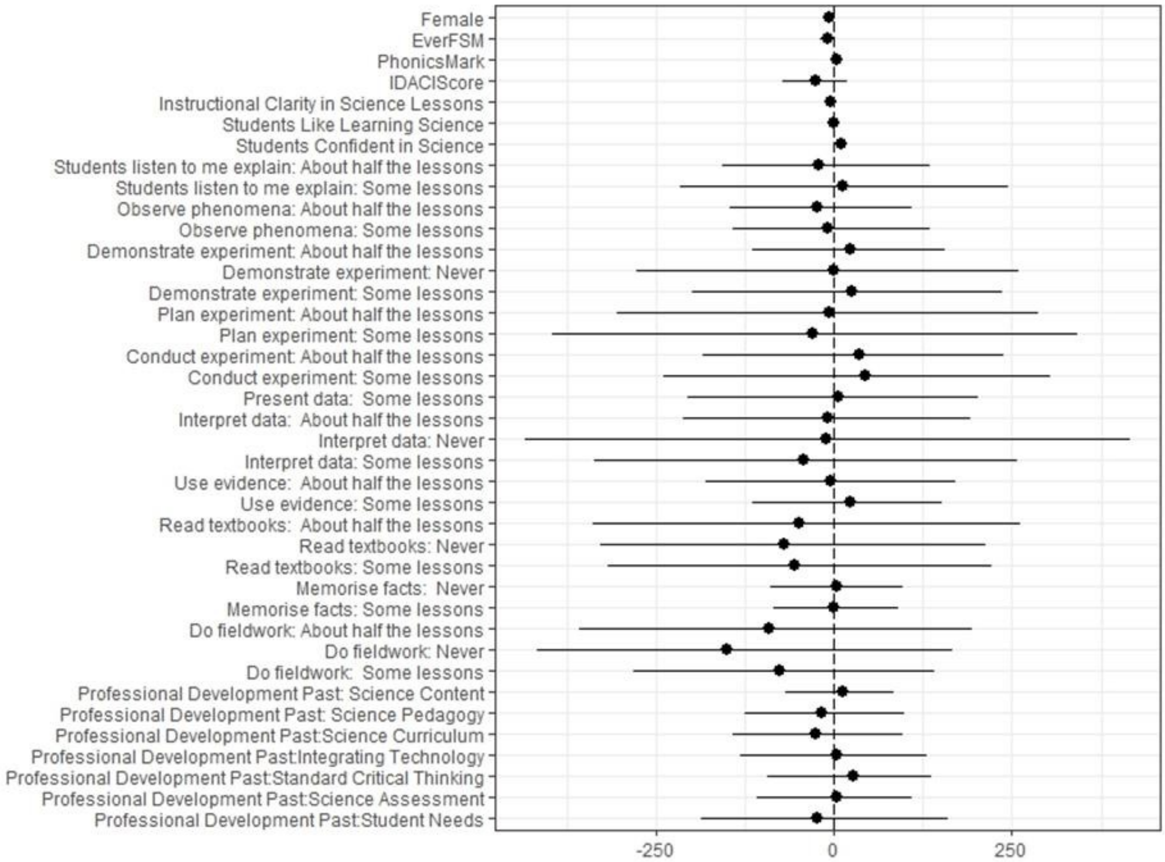
As described earlier, a model was developed that drew from TIMSS and the NPD. To summarise, we modelled the pupil outcome on the TIMSS overall science score (using the first plausible value) against:

- three demographic measures: sex of the pupil, the free school meal entitlement and IDACI score, and one measure of prior achievement, the score on the phonics screening check in year 1;
- three measures capturing the pupil experience of learning science: how clear their teaching has been, their enjoyment of science and their confidence in learning science;
- teachers' teaching strategies and experiences of CPD.

Furthermore, we included multilevel adjustments (varying intercepts) for class and school attended, to control for the inherent clustering of pupils in these groupings. Missing data was a considerable issue, and more than half the teachers within this sample (England) did not engage with the survey at all. Of those that did engage with the teacher survey, even fewer engaged with the question on attendance of science related professional development courses. The probability of observation is presented in the table below. Given the complexity of the multilevel structure and extent of missing data, there were few options available to impute the missing values. Our investigation found that the Markov Chain Monte Carlo (MCMC) simulations algorithms did not converge in sufficient time to allow an imputed analysis to be conducted.

The coefficient plot in figure 5 shows the pupil and teacher level variable coefficients from our final model. There were 658 pupils clustered in 33 classes and 33 schools compared to the original sample of 3319 pupils clustered in 159 classes and 138 schools. Given the extent of the missing data which resulted in 80% missing cases the teacher level variable estimates all crossed 0 with extremely wide intervals.

Figure 5. A coefficient plot of variables and pupil achievement



From the results of this model, and focusing on those parameters that were or close to traditional levels of significance, the average science score was 648.4 (252.0, 999.9) with females on average scoring 6.4 points lower than their male counterparts (-15.6, 2.7) although the interval crossed 0. Free school meal eligible pupils on average scored 9.0 points lower (-20.2, 2.2) although again the interval crossed 0. Prior achievement is a predictor: for the score on the phonics test, a one-unit improvement led to an increased score on the science assessment of 3.2 points (2.6, 3.8). For the IDACI score, a unit change from lower levels of deprivation to higher levels of deprivation saw an average reduction in the score of -26.5 points (-71.2, 18.7). The level of deprivation has a much larger impact on pupil achievement than any other variable we have considered.

Higher levels of instructional clarity as reported by pupils, were associated with lower average scores, with a one-unit change resulting in an average score 4.6 points lower (-7.5, -1.9). This appears to be counter intuitive as a finding and not in line with international data. The official TIMSS analysis (Mullis et al., 2020) reported that higher levels of instructional clarity are associated with higher levels of achievement. This is possibly due to low teacher participation of the sample of teachers in England, and our model includes teachers as well as pupil data, which has impacted both the available data and the analysis. Given that missingness is substantial and potentially missing not at random we are extremely cautious about the results.

Higher levels of liking science were associated with a very minor reduction in the average score by -0.5 points although the interval crossed 0 (-3.3, 2.3). Lastly, higher levels of confidence in science were positively associated with the science score with a one-unit change resulting in a 10.4-point increase (6.8, 13.9). From these it appears that supporting pupils in learning science needs to include tending to the affective, where pupils can be confident in their ability to learn science and enjoy the experience of learning science. Internationally, too, pupils who reported liking science, had higher average achievement scores.

With respect to teachers, the data analysis shows wide intervals with a small number of exceptions. Asking pupils to memorise facts seems to have little impact on achievement, more frequent teacher explaining is positively associated with achievement and certain and specific aspects of investigative or experimental work associated with achievement: observing phenomena, asking pupils to plan experiments and interpret data more frequently seem to be associated with higher levels of achievement whereas conducting experiments in all science lessons seems to be less effective. Indeed, working scientifically is developmental, iterative and needs to be supported (Luxton & Pritchard, 2023). The use of textbooks in primary science in England is not well documented and in countries where this is more common, it is associated with higher levels of attainment.

The survey questions on teacher professional development (PD) were towards the end of the questionnaire and were relatively poorly answered by teachers in England. The Improving Primary Science Guidance Report (Luxton & Pritchard, 2023) highlights the importance of effective evidenced-informed professional development. As noted earlier in this report, teachers who did answer these questions identified rather different aspects as the focus for their future professional development. For example, the teachers in the TIMSS survey identified (from a pre-populated list) PD needs in respect of integrating science with other subjects, integrating technology in the curriculum, pupil needs, science assessment, critical thinking as priorities with less emphasis on science content, science curriculum and science pedagogy. The EEF Guidance report (Luxton & Pritchard, 2023) includes evidence-informed and practical recommendations for primary schools in this country to include the development of teachers through professional development programmes.

Conclusions

The low participation rate and incomplete responses by teachers in England suggests we may not have a country-wide representative population. In turn, this limits the lessons learned or the understanding of primary science teaching and learning. Given these caveats, from the analysis of the data we have been able to use, the following conclusions can be drawn:

1. **Instructional time and clarity**
   We have not found an association between instructional time or instructional clarity and achievement for pupils in England. In comparison, analysis of the international data shows that there is a 'sweet spot' of instructional time, so efforts to supplement science learning experiences at this level of schooling with additional out of school teaching time is not beneficial. The TIMSS report details that, 'internationally and within most countries, however, higher clarity was associated with higher average achievement' (Mullis, Martin et al., 2019).

2. **Frequency of investigation, pupil enjoyment and interest in science**
   We have found that specific aspects of classroom scientific investigative work are positively associated with achievement. When pupils reported greater frequency of teachers asking questions to assess their learning, this was positively associated with attainment. This augurs well for supporting teachers to develop a 'thinking classroom' and as recommended by the EEF Primary Science Guidance Report, particularly in 'cultivating reasoning and justification'. By contrast, emphasis on the procedural (planning, conducting) with frequent use of classroom experimental work was negatively associated with attainment.

3. **Teachers**
   TIMSS data showed that the highest achievement for Year Five pupils was associated with teachers in England who had a background of education in science.

Higher levels of attainment are also associated with pupils:

- being confident in learning science;
- being asked to interpret data and use evidence;
- reading textbooks and source materials;

We note that that the phonics score at the end of Year One is closely correlated with the Year Five TIMSS score, that KS2 data was unavailable and that the sampling science tests for KS2 were not undertaken in this year.

4. Professional development experiences: past and future

Limited participation by teachers in England of the TIMSS teacher survey limits the lessons that can be learned from the analysis. In sum, we have not established a causal relationship between PD for elementary (primary) teachers and their pupils' attainment. The 'gap' in experience and future needs for CPD are not so much concerned with science content or curriculum but rather more general including the integration of science, the integration of technology, and meeting pupil needs.

Concluding remarks

However, the aspects that pupils bring to school themselves, the influence of their home background is much larger than any of the above factors. The IDACI score has a much larger effect on pupil achievement where this measure of deprivation is reflected in lower pupil achievement scores. As reported earlier this year (EPI, 2024), the attainment gap exists in early childhood and continued to widen throughout their schooling. This suggests that very early interventions to ameliorate the effect of economic disadvantage might result in more equitable attainment across pupils.

References

Bianchi, L., Whittaker, C., & Poole, A. (2021). *The 10 key issues with children's learning in primary science in England.* The University of Manchester and The Ogden Trust. Retrieved from https://documents.manchester.ac.uk/display.aspx?DocID=57599.

CFE. (2019). *Understanding the 'state of the nation' report of UK primary science education*: A baseline report for the Wellcome Trust. Leicester. Retrieved from https://cms.wellcome.org/sites/default/files/understanding-state-of-the-nation-report-of-uk-primary-science-education.pdf.

Darling-Hammond, L., Hyler, M. E., & Gardner, M. (2017). *Effective Teacher Professional Development.* Retrieved from Palo Alto, CA: https://learningpolicyinstitute.org/product/teacher-prof-dev

Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher, 38*(3), 181–199. https://doi.org/10.3102/0013189X08331140

de Vries, J. A., Dimosthenous, A., Schildkamp, K., & Visscher, A. J. (2022). The impact on pupil achievement of an assessment for learning teacher professional development program. *Studies In Educational Evaluation, 74*, 101184. doi:https://doi.org/10.1016/j.stueduc.2022.101184

Diamond, B. S., Maerten-Rivera, J., Rohrer, R. E., & Lee, O. (2014). Effectiveness of a curricular and professional development intervention at improving elementary teachers' science content knowledge and pupil achievement outcomes: Year 1 results. *Journal of Research in Science Teaching, 51*(5), 635-658. doi:10.1002/tea.21148.

Education Policy Institute. (2024). Annual Report

Hattie, J. (2009). *Visible Learning: A Synthesis of Over 800 Meta-Analyses Relating to Achievement.* London, UK: Routledge.

Jerrim, J., Sims, S., & Oliver, M. (2023). Teacher self-efficacy and pupil achievement: much ado about nothing? International evidence from TIMSS. *Teachers and Teaching, 29*(2), 220–240. https://doi.org/10.1080/13540602.2022.2159365

Luxton, K., & Pritchard, B. (2023). *Improving Primary Science Guidance Report*. Retrieved from London:

Maeng, J. L., Whitworth, B. A., Bell, R. L., & Sterling, D. R. (2020). The effect of professional development on elementary science teachers' understanding, confidence, and classroom implementation of reform-based science instruction. *Science Education, 104*(2), 326-353. doi:https://doi.org/10.1002/sce.21562.

Martin, M. O., von Davier, M., & Mullis, I. V. S. (Eds.). (2020). *Methods and Procedures: TIMSS 2019 Technical Report*. Retrieved from Boston College, TIMSS & PIRLS International Study Center website: https://timssandpirls.bc.edu/timss2019/methods

Mullis, I.V.S., Martin, M. O.,Foy, P., Kelly, D.L. & Fishbein, B. (2020). *TIMSS 2019 International Results in Mathematics and Science.* International Centre for Evaluation of Educational Achievement, Retrieved from https://www.iea.nl/sites/default/files/2020-12/TIMSS-2019-International-Results-in-Mathematics-and-Science.pdf.

Mutch-Jones, K., Hicks, J., & Sorge, B. (2022). Elementary science professional development to impact learning across the curriculum. *Teaching and Teacher Education, 112*, 103625. doi:https://doi.org/10.1016/j.tate.2021.103625.

Office for Standards in Education. (2019). *Intention and substance: further findings from the phase 3 research: primary science*. Retrieved from https://assets.publishing.service.gov.uk/media/5fb3ea67d3bf7f63def3669b/Intention_and_substance_findings_paper_on_primary_school_science_110219.pdf .

Ofsted. (2023). Finding the optimum: the science subject report. Retrieved from https://www.gov.uk/government/publications/subject-report-series-science/finding-the-optimum-the-science-subject-report--2.

Oliver, M., McConney, A. & Woods-McConney, A. (2021). The Efficacy of Inquiry-Based Instruction in Science: a Comparative Analysis of Six Countries Using PISA 2015. *Res Sci Educ 51* (Suppl 2), 595–616. https://doi.org/10.1007/s11165-019-09901-0

Richardson, M., Isaacs, T., Barnes, I., Swensson, C., Wilkinson, D., & Golding, J. (2020). *Trends in International Mathematics and Science Study* (TIMSS) 2019: National report for

England. London, UK. Retrieved from https://assets.publishing.service.gov.uk/media/5fca467ad3bf7f5d09db26ae/TIMSS_2019_National_Report.pdf.
Sims, S., & Fletcher-Wood, H. (2021). Identifying the characteristics of effective teacher professional development: a critical review. *School Effectiveness and School Improvement, 32(*1), 47-63. doi:10.1080/09243453.2020.1772841
Stubberfield, L. (2021). Slowly does it – using research to improve primary science. *Journal of Emergent Science, (20),* 5-10.
Taylor, J. A., Roth, K., Wilson, C. D., Stuhlsatz, M. A. M., & Tipton, E. (2017). The Effect of an Analysis-of-Practice, Videocase-Based, Teacher Professional Development Program on Elementary Pupils' Science Achievement. *Journal of Research on Educational Effectiveness, 10(*2), 241-271. doi:10.1080/19345747.2016.1147628
Timperley, H. (2008). Teacher professional learning and development. Retrieved from Brussels:

Survey questionnaires for teachers, school leaders and pupils are available at TIMSS 2019 Questionnaires (Grade Four).