A SYSTEMATIC REVIEW OF APPROACHES TO PRIMARY SCIENCE TEACHING

November 2023

Judith M. Bennett[1]

Lynda Dunlop[1]

Lucy Atkinson[1]

Sarah Compton[1]

Helen Glasspoole-Bird[1]

Fred Lubben[1]

Michael J. Reiss[2]

Maria Turkenburg-van Diepen[1]

[1]University of York Science Education Group

[2]University College London Institute of Education

The EEF Guidance Report is available at:
https://educationendowmentfoundation.org.uk/education-evidence/guidance-reports/primary-science-ks1-ks2

Co-funded by Wellcome

The Education Endowment Foundation (EEF) is an independent grant-making charity dedicated to breaking the link between family income and educational achievement, ensuring that children from all backgrounds can fulfil their potential and make the most of their talents.

The EEF aims to raise the attainment of children facing disadvantage by:

- identifying promising educational innovations that address the needs of disadvantaged children in primary and secondary schools in England;
- evaluating these innovations to extend and secure the evidence on what works and can be made to work at scale; and
- encouraging schools, government, charities, and others to apply evidence and adopt innovations found to be effective.

The EEF was established in 2011 by the Sutton Trust as lead charity in partnership with Impetus Trust (now part of Impetus - Private Equity Foundation) and received a founding £125m grant from the Department for Education. Together, the EEF and Sutton Trust are the government-designated What Works Centre for improving education outcomes for school-aged children.

For more information about the EEF or this report please contact:

Education Endowment Foundation
5th Floor, Millbank Tower
21–24 Millbank
SW1P 4QP

0207 802 1653

info@eefoundation.org.uk

www.educationendowmentfoundation.org.uk

Table of Contents

Acknowledgements

We are grateful for the support from the team at the Education Endowment Foundation: Bob Pritchard, Hannah Shackleton, Jonathan Kay, Nikki Kaiser, Rupal Patel, and in particular to Katie Luxton for the constructive conversations throughout the review.

We would like to thank the guidance panel for their feedback at critical points during the review: Alison Eley, Amy Halsall, Ben Rogers, Jane Turner, Lynne Bianchi, Louise Parks, Marianne Cutler, Nadia Moustapha and Sarah Earle. We are also grateful for the support of Steve Higgins and colleagues, and of the teachers who participated anonymously in the review of practice.

Lastly, we would like to acknowledge and thank Wellcome, who co-funded the evidence review with the EEF.

Executive summary

Introduction and background

Primary science is important for pupils as individuals and for society: amongst many other reasons, it can help children to understand and reason about themselves and the world, enable them to live healthy lives, and make informed choices (Harlen, 2018). It is important for society to have a scientifically literate population to respond to major social and environmental threats to human wellbeing such as climate change, food supply, pandemics and energy production (Royal Society, 2010). Concerns have been expressed by Ofsted (2023) amongst others that the status of science in primary schools has been lower since the removal of national tests in 2009, with pupils in some schools going an entire half term without learning science. England's position in the grade 4 (age 10-11) TIMSS (Trends in International Mathematics and Science Study) comparative tests, whilst above the average of participating countries, is outside the top ten (Mullis *et al.*, 2020). It is therefore important to understand how to best organise, teach and assess primary science. Whilst several reviews have focused on identifying the effectiveness of specific approaches such as inquiry (Slavin et al., 2014), the purpose of this study was to review the effectiveness of different approaches to primary science teaching on a range of pupil outcomes.

Approach and methods

Informed by a review of practice to identify outcomes important to teachers, we systematically searched electronic bibliographic databases and other educational sources for studies reporting on the effects of different approaches to primary science curriculum, teaching and assessment, published and including data collected between January 2007 and September 2021. A total of 15,476 studies were screened. We included studies if they had a sample of children aged between 5 and 11 and reported a science-specific teaching approach or a general teaching approach used in a primary science context. Studies were included if they had a counterfactual and where they reported pupil-level outcomes. The final body of literature for review included 90 studies.

The majority of studies used quasi-experiments (66 studies), with a smaller number of randomised control trials (23 studies) and one randomised field trial. The majority of studies (77) involved children aged 8 or above, with only 13 including children under 8 in the sample. Studies were conducted in a range of countries, with over half carried out in the USA (36) and Taiwan (16), and the remainder elsewhere. Only 6 studies were conducted in the UK. Studies were assigned to one of nine clusters determined by the review team in consultation with the EEF and Guidance Panel: assessment and feedback; context-based and cross-curricular approaches; co-operative and collaborative approaches; critical thinking and argumentation; explicit instruction and related approaches; ICT-supported and online teaching and learning; language, literacy and text-based approaches; learning outside the classroom; and practical work, inquiry and investigation.

In each cluster, we identified the constituent approaches, nature of the comparison group, outcomes measured and the summary of effects, along with identifying requirements for implementation, gaps in the evidence base and applicability of the evidence base to the English context. We found a diverse range of approaches, desired outcomes and measures across many different primary science topics and as a result, it was not meaningful to conduct a meta-analysis or to rank approaches. Few approaches were independently evaluated or used existing research instruments.

Findings and implications

We found a relatively large number of studies, multiple high-quality studies, with large sample sizes. Positive effects were reported across a range of outcomes including attainment and scientific language. Approaches included dialogic teaching, science and language integrated instruction and science for language learners. We found evidence to suggest that integration of science and literacy over an extended period can improve cognitive and affective outcomes for pupils in science *and* literacy, particularly for pupils who are less proficient in the language.

A connection between language, thinking and science was common in a number of approaches across clusters, indicating that increased talk enables pupils to identify prior knowledge and evidence to support their thinking, integrate new ideas, practise articulating these and receive feedback from peers and teachers. Relatedly, a number of studies involving cooperation and collaboration reported positive effects on attainment and attitudes. These included collaborative writing, discussion, online collaboration, and collaboration during differentiation.

Few studies reported on the effectiveness of different approaches to assessment, but those reporting on formative assessment approaches which create opportunities for meaningful classroom interactions focused on the learning goal, and time for thinking and talking (embedded formative assessment, guided peer feedback, bidirectional peer assessment) reported positive effects.

Several studies across clusters pointed to positive effects on attainment of giving children opportunity to learn outside the classroom, whether linked to school grounds and local habitats or shorter-term field visits to science centres, planetaria, or nature parks.

Few studies reported differential outcomes for different groups of pupils. For those eligible for free or reduced school meals, approaches which provide opportunities to make thinking explicit through words, whether spoken or written, seem to show promise, for example, dialogic teaching and the use of writing for argumentation. For pupils with low prior attainment compared with peers, approaches with additional benefits included those involving thinking time and classroom talk. Approaches which reported specific benefits for pupils who speak English as an additional language include scaffolded instructional discourse and instruction which integrates scientific content with literacy instruction. Finally, approaches reporting gains for children with special educational needs included guided inquiry involving activation of prior knowledge, explicit instruction of vocabulary, read-alouds and investigation activities. A frequent moderating factor across approaches was the provision of CPD for teachers, and whole school support for the approach.

Gaps in research evidence exist in relation to approaches that are used by teachers and those that have been investigated using research designs with a counterfactual. The review points to the need for a research agenda aligned to the aims of primary science education, and which is dedicated to accumulating an evidence base on approaches to the curriculum, teaching and assessment.

Limitations

The review included studies published and containing data collected between 2007 and 2021 and published in English, meaning that high-quality studies of approaches in areas of research that have been less popular in recent decades, and those published in languages other than English are not included. The predominance of small-scale studies reporting statistically significant positive results of approaches used in interventions suggests some publication bias. Finally, the review did not include studies that focused only on the early years foundation stage, so the review will have limited applicability to teachers of the Reception year in primary schools in England.

Part A: Aim, background and methods

A1 The aim of the review

The principal aim of this systematic review is to synthesise the evidence on effective teaching of primary science in schools. Specifically, the review addresses the question, *'What approaches are most effective to improve pupil outcomes in primary science, in what context, and how?'*.

For the purposes of this report, the term *primary science* is used to refer to teaching and learning associated with the curriculum subject *Science* for pupils aged 5-11. This is the equivalent of Key Stages 1 and 2 in England. Studies involving children only in the early years are excluded from the review because the early years foundation stage (EYFS) is often subject to different aims, priorities, areas of learning and outcomes. However, the studies included in this review are international, and there is variation from country to country in the age at which children start their primary or elementary education. Therefore, some of the studies in the review include children under the age of five, but these have only been included if they also include children aged five and over.

In England, primary science education involves teaching and learning concepts across biology, chemistry, and physics, as well as how science works: the nature, processes and methods of science. Within this, there is the need to learn new subject-specific vocabulary, new uses for everyday vocabulary, and mathematical skills such as pattern seeking and data handling. Children also learn to think about knowledge, how it can be applied, and the implications of science for today and tomorrow (DfE, 2013). This review focuses on effectiveness in relation to a range of outcomes identified as important by teachers including attainment, conceptual knowledge and understanding and attitudes towards science.

A2 Background and review rationale

Scientific, policy and practical background

A number of reviews[1] relating to primary science education have been published over the past ten years. These tend to cover both primary and secondary phases (Hartmeyer *et al.,*2018; Nunes *et al.*, 2017; Potvin & Hasni, 2014) and come from the perspective of understanding subject-specific concepts (Lelliott, & Rollnick, 2010), assessment approaches (Hartmeyer *et al.*, 2018), science aspirations (DeWitt & Archer, 2015; Dewitt *et al*., 2014) and specific teaching approaches (Huerta & Garza 2019) or the impact of broader contextual factors which have an impact on science education across phases (Banerjee, 2016). One existing review of primary science teaching (Gresnigt *et al.*, 2014) examines the effectiveness of different approaches to integrating science across the primary curriculum, but not on different approaches to teaching primary science. Another review (Slavin *et al.*, 2014) identified approaches to primary science based on stringent inclusion criteria, such as having a control group, intervention duration of at least four weeks, and using outcome measures not inherent to the experimental treatment (amongst other inclusion criteria). Prior to the present study there has not yet been a comprehensive systematic review of effective approaches to teaching primary science.

[1] These include both systematic and narrative reviews. A systematic review uses transparent procedures to systematically find, evaluate and synthesize the results of relevant research. The procedures are often explicitly defined in advance, to ensure that the exercise can be replicated and to minimise bias (Campbell Collaboration). A narrative review summarises evidence on a specific topic of interest but may or may not use systematic procedures.

Why the effective teaching of science in primary schools is important

Society is facing huge challenges today, including climate change (IPCC, 2023), pandemic-level disease threats (Wellcome, 2022), antibiotic resistance (WHO, 2020), pollution (Fuller *et al.*, 2022) and biodiversity loss (UNEP, Convention on Biological Diversity and WHO, 2015). Science is one of the key means of helping society address these challenges. As many have argued (see, for example, Bybee, 1997; Millar and Osborne, 1998; Millar & Osborne, 2006; Norris & Phillips, 2003) it is therefore essential that young people are provided with the scientific knowledge, skills and abilities they need to understand and act in the world around them – in other words, they need to be scientifically literate citizens, whether or not they go on to be future scientists. To enable this to happen, young people need a science education that enables them to understand and appreciate what science is, how scientific knowledge is created and used, and how science has an impact on individuals and society.

Harlen (2018) suggests that the reasons for the importance of science education can be divided into three groups: for learners as individuals, for society, and for learners as global citizens. For individuals, science education helps them lead healthy and rewarding lives, stimulates curiosity, develops skills needed for a rapidly changing world, and to make informed decisions about the role of scientific evidence. For society, science helps people make informed choices about matters such as the use of energy and resources, ensure that scientific knowledge is used appropriately, be aware that the application of scientific knowledge through technology can have both positive and negative impacts on society, and develop informed attitudes to the ways in which science can be used. For global citizenship, science can help respond to global challenges and answer questions such as how to ensure food and water for all, how to tackle disease, and what is needed in response to the climate emergency.

Primary science as a compulsory part of the school curriculum

The arguments for the importance of science apply to all young people. Whilst science subjects have had a secure place in the secondary school curriculum for several decades, this has not been the case in the primary school curriculum.

The need to include science in the primary school curriculum in England and Wales was first advocated in the 1960s, partly as a possible way of increasing the numbers of scientists. It was also strongly advocated by UNESCO (the United Nations Educational, Scientific and Cultural Organisation) in the early 1980s (UNESCO, 1982). The first moves towards making science compulsory in primary schools in England and Wales came in 1985, when the Department of Education and Science and Welsh Office published a policy statement saying that all pupils should be introduced to science in primary schools (DES and WO, 1985). One outcome of this was that, when the National Curriculum for ages 5-16 was introduced in England and Wales in 1988, English, mathematics and science were specified as the three core subjects all pupils had to study. In addition, at primary level, pupils had to sit Standard Assessment Tests (SATS) at ages 7 and 11 in these core subjects.

Research into the teaching of science in primary schools

At the time of the introduction of the National Curriculum, there was very little research into the teaching of science in primary schools. One exception to this was the work done by the Government-funded Assessment of Performance Unit (APU) in the first half of the 1980s.

The APU was established to undertake annual surveys of English, mathematics and science for pupils at ages 11, 13, and 15 in England, Wales and Northern Ireland. The inclusion of age 11, the last year of primary school, was seen as an attempt to give more status to science in primary schools (e.g. Harlen, 2018). The work of the APU was very influential, as it gave rise to one of the most-widely researched areas in science education, that of pupils' understanding of key ideas in science (variously known as misconceptions, alternative frameworks, and children's ideas).

The Children's Learning in Science Project (CLIS) project, which ran from 1984-1986, undertook much of the early work, which focused on secondary school pupils. The project revealed a number of very significant areas where secondary pupils' understanding of science ideas differed from accepted understanding in science (e.g. forces, electricity, light, plant growth, inheritance, particulate nature of matter). The work was extended to primary age children through the Science Processes and Concepts Exploration (SPACE) project, which ran from 1990 to 1998, and revealed a range of ideas about science that primary age children had developed trying to make sense of their experience of life. As with their secondary counterparts, not all these ideas matched with accepted scientific understanding, but proved to be firmly held. These studies gave rise to the constructivist approach to learning in primary science (Driver, 1983), where the ideas that children construct for themselves to explain their environment are not ignored but used as the starting point for developing scientific understanding.

Since the introduction of the National Curriculum, research into aspects of the teaching of primary science has, not unexpectedly, become much more common. Areas of research focus on curriculum (Sharpe & Grace, 2004), the effect of different approaches (e.g. inquiry learning (Slavin *et al.*, 2014), collaborative learning (Tolmie et al., 2010), practical work (Abrahams & Reiss, 2012), the use of real-life contexts in teaching (e.g. Guerra-Ramos *et al*., 2010), assessment (Earle, 2014), the use of ICT (Murphy & Beggs, 2003), teacher confidence (Murphy *et al*., 2008) and the impact of CPD courses on teachers and their pupils (Bennett *et al.*, 2019).

How has science fared as a core curriculum subject in primary schools?

While science becoming a core curriculum subject in primary schools was seen by most people as a positive move, it created a number of challenges. These relate to the nature of the curriculum, assessment, approaches to teaching, and equipping primary school teachers with the knowledge and skills to teach science.

The nature of the curriculum

When it was originally introduced, the National Curriculum for Science specified a range of content to be taught at four key stages: Key Stage 1 (age 5-7), Key Stage 2 (age 7-11), Key Stage 3 (age 11-14) and Key Stage 4 (age 14-16) (DfE & QCA, 1999). In addition to subject knowledge, a strand on scientific inquiry was also introduced. What was specified at each key stage was largely driven by a consensus approach, rather than by research findings.

There have been a number of changes to the science curriculum since it was introduced in 1988. Two of particular note were the supplementing of the National Curriculum with Schemes of Work for each topic in the late 1990s, and modifications to the National Curriculum to provide a more knowledge-rich curriculum in 2014. There have also been changes of emphasis in relation to the process of developing scientific knowledge, which has variously been referred to as 'scientific inquiry' (pre-2000), 'how science works' (post-2000) and 'working scientifically' (post-2013). Recent reforms mean that academies are not required to follow the National Curriculum, although they must provide a broad and balanced curriculum.

With the exception of the Schemes of Work, the science curriculum has not been supported by any Government-led resources. At primary level, the last thirty years has therefore seen a number of curriculum resources being developed, by a range of groups: publishers, subject associations, private companies, funding bodies and universities. Some of the materials developed cover the whole curriculum and others focus on ideas for practical work. Appendix 1 to this report describes a number of these resources. Thus, teachers have a range of resources on which they can draw. However, with a small number of exceptions, few of these resources have been developed with explicit reference to research literature.

Alongside the 'consensus approach' to what should be in the curriculum has been the notion of 'big ideas' in science. Harlen *et al.* (2010) outline 'principles and big ideas of science education.' Rather than conceiving science education as a body of facts and theories – which often appear disconnected to learners – it should be seen as a progression towards a small number of 'big ideas' that, together, enable understanding of events and phenomena of relevance to pupils' lives during and beyond their school years. Harlen *et al.* (2020) went on to specify fourteen big ideas in science to which young people should be introduced. Examples of these big ideas include: All material in the Universe is made of very small particles, The total amount of energy in the Universe is always the same but energy can be transformed when things change or are made to happen, The diversity of organisms, living and extinct, is the result of evolution. The notion of 'big ideas' as a unifying structure for the curriculum is seen as very attractive, and work is currently being done by the Learned Societies (Royal Society of Biology, Royal Society of Chemistry, Institute of Physics) to try and ensure any future version of the National Curriculum takes account of 'big ideas' (see e.g. McLeod, 2018). These 'big ideas' are based on principles that should underpin the science education of all pupils, based on the ideas of a group of international experts in science and science education during a seminar supported by the 'PuRkwa' prize, given by the French Academy of Sciences and the Saint Etienne Mining School (Harlen, 2010).

Assessment

Arguably, the most substantial impact on the science experiences young people have had in primary schools has come from assessment and monitoring practices. By the mid-1990s, the Government was becoming increasingly concerned about SAT (Standard Assessment Test) targets not being met in English and mathematics and, in 1998-9, introduced national frameworks for teaching literacy and numeracy (DfEE, 1998; DfEE, 1999). The effect of this was to downgrade the importance of other subjects, including science (Boyle & Bragg, 2006) though science still remained, as it does today, a core subject.

A further challenge for science in the primary curriculum arose from the impact of assessment. In 2009, in response to widely held perceptions of an excessive testing culture and too much emphasis on 'teaching to the test' (e.g. Murphy, Neil & Beggs, 2003; Collins, Reiss & Stobart, 2010), the Government abolished science SATs in primary schools in England. This was seen as a change that would allow teachers to be more creative in their science lessons and less preoccupied with assessed outcomes (Harlen, 2018). However, by 2014, the Wellcome Trust (Wellcome Trust, 2014) reported that the removal of SATs, coupled with greater emphasis on outcomes for English and mathematics, had led to a diminished status for science in the primary curriculum, and a substantial reduction in the time allocated to science in many primary schools. This tension between the desire to provide all children with a good background in science in primary schools, without the curriculum being influenced by the potential negative effects of assessment, persists to the present day.

Of relevance to assessment are the international and national studies of pupil achievement. These are large-scale assessments of a sample of students that seek to measure performance and look at trends over a period of time. The Trends In International Mathematics and Science Study (TIMMS) has been reporting international data every four years on pupils aged 10. More recently, in 2014, the Standards and Testing Agency in England introduced its own national system of sampling to monitor performance in science of students aged 10-11 every two years. While these data are unlikely to directly influence the way in which teachers approach assessment, they do serve to inform national decisions on policy. The TIMSS data suggests that performance in England improved between 1995 and 2007, declining sharply in 2011, with some recovery since that time (Mullis *et al.,* 2020). This may be linked to changes in the National Curriculum and in assessment. Thus far, there has been no significant difference from year to year in the STA assessments (STA, 2019), but these were introduced after the removal of SATs at age 11.

Approaches to teaching

A wide diversity of approaches to teaching is apparent in the resources developed to support teachers in their teaching of primary science. These include inquiry-based approaches, practical work, problem-solving, the role of talk, discussion and questioning, development of thinking skills, self-regulated learning, explicit instruction, and creative approaches including drama and role-play.

Inquiry-based approaches are common in science teaching. Harlen (2013) notes that inquiry can be applied to a number of curriculum subjects but that 'what distinguishes scientific inquiry is that it leads to knowledge and understanding of the natural and made world through direct interaction with the world and through the generation and collection of data for use as evidence in supporting explanations of phenomena and events' (p.12). Inquiry-based approaches cover such aspects as fair testing, observation, looking for patterns, open-ended investigations and problem solving. Inquiry-based approaches often include practical work, which Abrahams *et al.* (2014, p.264) define as 'any type of science teaching and learning activity in which pupils, working either individually or in small groups are required to manipulate and/or observe real objects and materials.'

One of the debates about practical and inquiry work is the impact they have on pupils. Pupils often report enjoying practical work, but what science do they learn from it (e.g. Abrahams and Reiss, 2016)? This has resulted in the advocacy of a 'hands-on' and 'minds-on' approach to practical work (Abrahams *et al*., 2014) where the teacher deliberately plans for links between observations and scientific ideas. Similarly, McCrory (2017) describes how inquiry-based approaches can motivate and engage learners and that 'exploration and inquiry are crucial in developing process skills for children to construct their understanding of conceptual science,' (p8). Whilst harnessing curiosity and fostering a love of science through fun activities, McCrory notes that such approaches must be designed to promote progression of conceptual understanding, scientific thinking and reasoning skills and make explicit links between scientific concepts and the inquiry activity undertaken.

The 'Thinking, Doing, Talking Science' (TDTS) project promotes the use of a design for practical science that encourages children's learning and engagement through creative, interactive and cognitively challenging lessons. Wilson *et al*. (2018) emphasise that primary science activities can be simple but need to be effective. This can be accomplished through the development and employment of specific higher order thinking skills (based on frameworks such as Bloom's Taxonomy) and including dedicated time for discussion.

Explicit instruction (also called teacher-directed instruction) is the use of instructional approaches that are structured, sequenced, and led by teachers. It includes activities such as the teacher explaining a scientific idea to a class, a whole-class discussion led by the teacher, and the teacher demonstrating an idea. Direct instruction has been the focus of considerable interest in science teaching since the publication of the PISA (Programme or International Student Assessment) findings in 2016 (OECD, 2016) prompted debate about the relative merits of direct instruction and inquiry-based approaches. Direct instruction is also one of the approaches discussed in the 2021 Ofsted research review in Science (Ofsted, 2021), with their review noting that “analysis of pupil responses and outcome data from PISA 2015 reveals that teacher-directed science instruction is positively associated with science performance in almost all countries.” (p25). In relation to inquiry-based learning, the report notes the challenges for learning science through exploration when you are a novice learner with little prior knowledge, and the high cognitive load associated with searching for solutions. However, the issue is more finely grained than a simple consideration of direct teaching versus inquiry-based approaches: the nature of the direct instruction or the inquiry-based approach is important, as are what is most appropriate for the subject matter to be covered and the student being taught. In a study pre-dating the PISA 2015 survey, Cobern *et al*. (2010) compared the use of direct teaching with inquiry-based approaches. Their findings led them to conclude that, while a choice between direct teaching and inquiry-based approaches may be appropriate for particular aspects of science, “good direct and inquiry instruction led to similar understanding of science concepts and principles in comparable times” (p93). Mourshed *et al*. (2017, p8) used the PISA data to identify what they refer to as “the sweet spot”, which is the use of teacher-directed instruction in most to all classes and inquiry-based learning in some. (See Figure 1.)

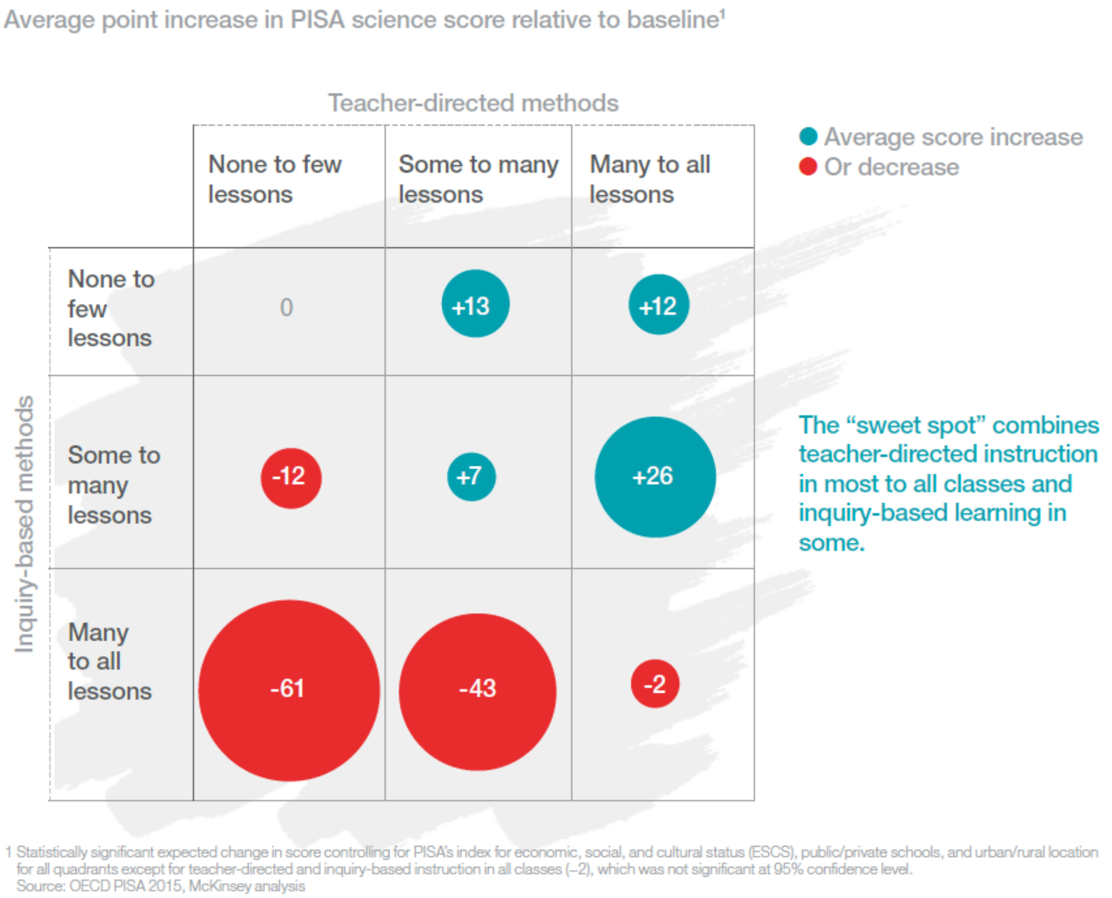


Figure 1: The ‘sweet spot’ for the use of direct instruction and inquiry-based approaches

Equipping primary school teachers with the knowledge and skills needed to teach science

A substantial challenge faced when the National Curriculum was introduced was equipping primary school teachers, most of whom were not science specialists, with the knowledge, skills and resources needed to teach science. Thus, the late 1980s and 1990s saw considerable growth in continuing professional development (CPD) courses for primary school teachers to help with teaching science, a trend that has continued into the 2000s. The need for CPD is illustrated by the range of organisations including but by no means limited to: the Association of Science Education (ASE), The Ogden Trust, the Primary Science Quality Mark (PSQM), Primary Science Teaching Trust (PSTT), and STEM Learning. These groups include professional associations, charity trusts, national training centres and higher education institutions.

Despite the wide range of provision, other pressures on teachers mean that creating the time for CPD is problematic, and issues relating to subject knowledge and confidence persist. A baseline survey on the teaching of science in primary schools commissioned by Wellcome (Leonardi *et al.*, 2017) noted that 25% of teachers were concerned that they may not be able to answer children's questions about science. Similarly, Mackintosh *et al.* (2017) found that most primary school teachers expressed a lack of confidence in their abilities to teach science. A recent cost-benefit analysis of CPD for teachers commissioned by Wellcome (Van den Brande & Zuccollo, 2021) found that primary teachers in England surveyed spend, on average, 55 hours a year on professional development. This is slightly less than the OECD average of 62 hours a year, and there are concerns over the quality of some of this CPD. The report attempts to quantify the benefits of CPD. These benefits include increased lifetime earnings for pupils and improved teacher retention.

In contrast to secondary schools, specialist science teachers are rare in primary schools with the Department for Education estimating that only 5% of primary teachers hold a science degree or specialised teaching qualification (DfE, 2021), so teachers taking on leadership roles are unlikely to have science qualifications above the level of GCSE or A-level. However, they will be responsible for leading the development of science teaching in their school, including training and supporting their colleagues.

Rushton and Reiss (2021) conducted a systematic review of the literature on science leadership in middle and high schools to explore the extent to which a social identity approach might provide a framework for structuring the findings of the literature on leadership. Their findings indicate that shared identity and group membership play a role in developing and sustaining leadership, and connecting with others from different schools appears to develop a shared role identity. In addition to CPD programmes, group membership was facilitated through networks and informal communities of practice formed using social media to connect with others in order to share knowledge, ideas and resources and keep up to date with government initiatives that impact on their role.

Some cautionary notes have emerged from the literature on the impact of CPD on science leadership. Bennett *et al.* (2019) undertook an evaluation of a national, large-scale CPD programme aimed at improving science leadership, including subject knowledge, in primary schools. The study findings indicated that the programme increased teachers' self-confidence, but there was no substantial impact on subject knowledge. In addition, teachers participating in the CPD very often focused on isolated activities to demonstrate to colleagues in their schools, rather than considering the associated subject knowledge.

Where are we now? Science in primary schools in the early 2020s

The most recent picture of the state of primary science education in England comes from the 'State of the nation' report, Evaluation of the primary science campaign, published by the Wellcome Trust (Wellcome, 2020), the Ofsted Research Review of Science (Ofsted, 2021), and the study by Bianchi *et al.* (2021) identifying ten key issues with children's learning in primary science in England.

Wellcome 'State of the nation' report

The Wellcome 'State of the nation' report (Wellcome, 2020) draws on a large-scale survey and interviews with science leaders in primary schools, identified through their use of the Explorify programme (a widely used website of free digital resources for teaching primary science).

The report identified a number of strategic issues in relation to the delivery of science in primary schools. Most schools had a dedicated staff member responsible for leading science development and teaching across their school, of which the majority were allocated time for their leadership role. Although science was believed to be important, it was seen as less important than English and mathematics, with one third of the schools surveyed not including science in their School Development Plan. Around half the teachers had undertaken externally provided CPD in the year prior to the survey. The use of Explorify resources was associated with a higher rate of participation in CPD, and greater in-school support for science.

In terms of provision, science was taught on average for 1.5 hours a week, with lessons often supplemented by other methods such as science weeks and science visits or other methods. Although most science leaders were positive about science and thought it was important for pupils to study the subject, many believed that not enough time was spent teaching science in their school. Within lessons, teachers often encouraged pupils to take part in class discussions and encouraged pupils to predict what will happen when they do science investigations.

The science leaders reported being confident in their ability to teach science, although they rated this lower than their confidence to teach English and Mathematics. Confidence in teaching science was linked to enjoyment of teaching science and to science being seen as important in a school.

Within the teacher survey, respondents were asked to state the first three words which came to mind when describing science, with 'Investigation', 'Experiment', and 'Practical' being the most frequently cited.

Ofsted Research Review of Science

The purpose of the Ofsted Research Review of Science (Ofsted, 2021) was "to identify factors that can contribute to high-quality school science curriculums, assessment, pedagogy and systems" (p3).

The review focuses on both primary and secondary science, and is structured around work in seven areas: Curriculum progression: what it means to get better at science, Organising knowledge within the subject curriculum, Other curricular considerations, Curriculum materials, Practical work, Pedagogy: teaching the curriculum, Assessment, and Systems at subject and school level.

In relation to the primary curriculum the review expresses concern about the status of science, and notes that "[t]he picture is not an improving one for all pupils and may be deteriorating" (p3). Drawing on the study by Bianchi *et al.* (2021), Ofsted notes that science for many pupils in primary schools consists of 'fun activities' without developing a deep understanding of the associated scientific concepts. Ofsted feels the situation with primary science is particularly concerning as experiences in primary schools have been shown to be crucial in shaping pupils' views of science and scientific aspirations (see, for example, the findings of the ASPIRES project in DeWitt & Archer, 2015).

In addition to noting the findings of the PISA 2015 survey on direct instruction, key recommendations from the Ofsted review of particular relevance to the teaching of primary science include:

- In primary schools, there is at least one teacher who specialises in teaching science and science leaders have dedicated leadership time.
- There is a need to sequence knowledge to promote effective learning.
- The curriculum is planned to build increasingly sophisticated knowledge of the products (substantive knowledge) and practices (disciplinary knowledge) of science.
- Substantive knowledge is sequenced so that pupils build their knowledge of important concepts such as photosynthesis, magnetism and substance throughout their time at school.
- The curriculum anticipates where pupils are likely to hold misconceptions and address misconceptions explicitly.
- Pupils need to learn about the different ways that scientists engage in their work through reading, writing, talking and representing science.
- The purpose of practical work is clear in relation to curriculum content so that practical activities can be set up and managed to develop pupils' disciplinary and/or substantive knowledge.
- Activities are carefully chosen so that they match specific curriculum intent.
- Teachers use systematic teaching approaches, where learning is scaffolded using carefully sequenced explanations, models, analogies and other representations to help pupils to acquire, organise and remember scientific knowledge.
- Systematic approaches, alongside carefully selected texts, are used to teach the most important vocabulary in science.
- Teachers and pupils are clear on the purpose of assessment.

The Ten Key Issues report by Bianchi et al. (2021)

The aim of the review by Bianchi *et al.* (2021) was to identify the issues which had an impact on children's learning in primary science. It drew on the knowledge and experience of a specialist team and a targeted survey of 72 wider stakeholders. As with the Wellcome report and the Ofsted review, it was underpinned by concern about the current state of teaching of science in primary schools, noting that "the profile of primary science has in recent years dwindled and it is frequently taught for fewer than the recommended hours with a reduced curriculum status than that which should be expected of a core subject. Primary teachers lack confidence and skills in science, which means that the sequencing of the curriculum is not always sensible and misconceptions are not corrected." (p3). A central feature of the report is the list of ten key issues in primary science (p6):

1. Children's science learning is superficial and lacks depth
2. Children's preconceptions are not adequately valued
3. Children's science learning lacks challenge
4. Children are over-reliant on teacher talk and direction, they lack autonomy and independence in learning science
5. Children experience 'fun' science activities that fail to deepen or develop new learning
6. Children are not encouraged to use their own curiosity, scientific interests and questions in their science learning
7. Children are engaged in prescriptive practical work that lacks purpose

8. Children do not draw on their learning from prior scientific skills, they do not build on repeated and regular experiences
9. Children rarely see themselves, their families, community members or their teachers as scientists
10. Children do not apply literacy and numeracy skills in science at the standard they use in English and mathematics.

Bianchi *et al.* (2021) strongly advocate the use of the report, and the 'Ten Issues' in particular, to focus dialogue within and beyond primary schools with a view to improving the quality of science teaching.

Practices in Primary Science

One way of considering practices in the teaching of primary science is to look at what is being advocated, and what teachers report doing in their lessons.

The Ofsted Review of Science (Ofsted 2021) is clearly one source of recommendations for practice (see above). Other sources include the work of Harlen (2018). Harlen recommends (p19) six key features of learning experiences in primary science. They should be:

- Engaging, interesting and relevant to children; motivating questioning and problem-solving
- Linking to children's existing ideas, building on previous experience and working towards big ideas
- Involving active investigation of real materials and events
- Promoting talk, dialogue and communication in various forms
- Using and developing inquiry skills
- Using technology to aid inquiry.

Looking at these two lists, one can see a number of similarities. Common suggested key features of good practice include the need to engage pupils, the need to explore existing ideas pupils may have before developing knowledge, the importance of sequencing ideas to build up a coherent picture of knowledge, and the importance of dialogue and communication. The role of well-structured practical work is also recognised, though Harlen's list places more emphasis on inquiry, as it does on the importance of 'big ideas' in science.

Existing narrative and systematic reviews

A number of reviews of relevance to primary science education have been published over the past ten years. The majority cover both primary and secondary phases, and explore the teaching of specific concepts, particular teaching approaches, attitudes and interest in science and broader contextual influences. Two additional reviews, the Ofsted Research Review of Science (Ofsted, 2021) and the Ten Key Issues report by Bianchi *et al.* (2021) have been discussed earlier in this report.

Reviews focusing on teaching approaches

Nunes *et al.* (2017) undertook a systematic review to gather and evaluate the evidence for promising educational approaches that are likely to improve the attainment and progression of low-Socio Economic Status (SES) pupils in science education. They concluded that interventions aimed at improving pupils' scientific reasoning or their scientific literacy have been generally successful because scientific reasoning and literacy are mediators for attainment in science. They also concluded that, although the number of studies was low, interventions designed to develop pupils' group work skills, and to teach them to evaluate and make use of their own assessment data were beneficial to their learning of science. They also found good evidence to suggest that residential fieldwork and the use of informal science education institutions improved pupils' learning of science, provided that these experiences were set up in carefully structured ways. There was some evidence to suggest that there were benefits of after-school activities, such as STEM clubs, and peer mentoring.

The Wellcome Trust (2017) conducted a review of the prevalence of different teaching approaches in primary science. This indicated that the most common approaches used in teaching were how to design and undertake investigations, demonstrating science activities, asking children to make predictions, record and interpret data and observations, and promoting participation in discussion.

Hartmeyer, Stevenson, and Bentsen (2018) conducted a systematic review of concept mapping-based interventions as a tool for formative assessment in primary and secondary science education, making a number of recommendations on how to use concept maps most effectively.

Huerta and Garza (2019) conducted a systematic review of interventions aimed at improving writing in science, with a specific focus on pupils whose first language is not English, and therefore face additional challenges in learning. They concluded that appropriately constructed writing tasks aid understanding in science.

Review focusing on understanding

Lelliott and Rollnick (2010) reviewed research into astronomy education, focusing on children's understanding of basic ideas about astronomy. They found that most studies focused on pupils' conceptions of ideas in astronomy, and that the field was now ready for research which focuses on using this evidence base to create teaching and learning sequences and resources, develop teachers' pedagogical content knowledge, and to design and test interventions to change pupils' conceptions. They also note the need for astronomy education research to better reach teachers in schools.

Reviews focusing on affective responses to science (interest, motivation, attitudes)

Potvin and Hasni (2014) reviewed research on interest, motivation and attitude towards science and technology, looking at the constructs employed and how they are defined, the instruments used, and the results presented. They note the widespread use of bespoke questionnaires rather than readily available instruments, and call for greater use of the latter. Potvin and Hasni (2014) also note the 'suspiciously high' number of experiments reporting positive results and suggest that the nature of the control condition may account for some of these positive results. They conclude that effort to increase interest, motivations and attitudes usually produces results, but that it is difficult to determine those of greatest benefit owing to the nature of studies in their review.

Reviews focusing on broader contextual factors

Banerjee (2016) conducted a systematic review to identify factors linked to the underachievement of disadvantaged pupils in school science and maths. 'Disadvantaged' related to factors such as socio-economic status (SES), barriers due to language, and ethnic background. Their results suggested that major factors linking deprivation to underachievement were linked to a lack of positive environment and support.

Slavin *et al.* (2014) conducted a systematic review to identify the effects of particular instructional approaches in primary science. The review found that programmes that used science kits (resources that contain lab equipment, materials, teaching manuals, alongside CPD) did not show positive outcomes on science achievement measures. On the other hand, inquiry-based programmes that emphasised professional development but not kits did show positive outcomes. They suggest that this may be linked to kits encouraging teachers to focus on activities at the expense of developing competence in strategies that can be used more generally to make science teaching engaging, comprehensible, and conceptually challenging. Approaches integrating video and computer resources with teaching and cooperative learning also showed positive outcomes. The review concluded that science teaching methods focused on enhancing teachers' classroom instruction throughout the year, such as cooperative learning and science-reading integration, as well as approaches that give teachers technology tools to enhance instruction, have significant potential to improve science learning.

It is worth noting that Slavin *et al. (2014)* concluded that "The most important finding of the present review is the very limited number of rigorous experimental evaluations of elementary science programs." (p878).

A3 Research question and objectives

Research question

The review addresses the question, 'What approaches are most effective to improve pupil outcomes in primary science, in what context, and how?'.

A systematic review was conducted to answer the research question. Systematic reviews allow researchers to characterise the quantity and quality of literature, to identify what is known and not known, and from this to generate recommendations for practice (Grant & Booth, 2009). The systematic review was informed by research on practice to contextualise the findings and ensure that the report focused on outcomes relevant to teachers. A theoretical review was conducted in parallel to understand the mechanisms by which successful interventions work.

We interpret 'approaches to science teaching' to include both science-specific approaches, such as practical work, and broader approaches, such as small-group work / cooperative learning, dialogic teaching, feedback, metacognition and self-regulation, where used in science teaching. 'Effectiveness' is defined in relation to attitudinal and attainment outcomes, as well as others deemed important to teachers during the review of practice (Appendix 2).

Strands of work

The review comprised three strands of work:

1. Empirical research into views of current practice
2. A theoretical review
3. A systematic review, including an evidence map.

Each strand of work was informed by the EEF guidance panel (see acknowledgements), an expert advisory group appointed and chaired by the EEF consisting of academics and practitioners in primary science. The guidance panel was invited to comment on documents and advise on key decisions throughout the review.

Empirical research into views of current practice

The empirical research into views of current practice was undertaken to identify current practices in order to inform decisions on outcomes of interest, and to provide contextual information. This would enable the findings of the systematic review to be presented in a way that was sensitive to the context of primary science education in England and help ensure the review was relatable and actionable for primary teachers.

The research on practice comprised two parts: (i) a narrative literature review focusing on practice in England and (ii) focus groups with primary teachers, subject leaders and headteachers. The literature review summarised recent research on practice, and the focus groups enabled questions to be asked that were pertinent to the review and reasons for primary science practices to be probed.

The following research questions were addressed in the research on practice:

- What outcomes are important to teachers of primary science?
- What approaches are currently used?
- How are current practices and perceptions aligned with research and theory on effective primary science teaching, prevalent pedagogical and curricular approaches, and the mechanisms that support science learning?

The narrative literature review consisted of a 'state of the art' review. State of the art reviews address more current matters in contrast to the combined retrospective and current approaches of a more general literature review (Grant & Booth, 2009) and are useful for identifying the current state of practice. A comprehensive search of literature published between October 2013 and August 2021 (i.e., from the introduction of the most recent version of the national curriculum in England, to present) was conducted. The review identified approaches that were being used, or being advocated for use, in the teaching of primary science (age 5-11) in England, together with any aims and outcomes associated with the approaches.

The report on the narrative literature review is found in Appendix 1.

Eight online focus groups were held with 31 teachers from 27 schools. Teachers were asked about the outcomes from primary science they thought were important, their approaches to teaching primary science, and influences on their teaching of primary science. Focus group recordings were transcribed and analysed using a framework developed by James and Brown (2007) to classify learning outcomes. Approaches to teaching were analysed inductively from the data and deductively using the Education Endowment Foundation's Improving Secondary Science guidance report (EEF, 2018) because some primary teachers have been reported to draw on these recommendations despite the differences in context, objectives and accountability methods in the primary sector (EEF, 2021). Transcripts were then analysed to identify additional approaches to and influences on primary science education. The review of practice was used to identify outcomes of interest for the systematic review and to inform key words for the search terms used to identify studies for inclusion. These were refined with input from the EEF, guidance panel and preliminary tests.

The report providing an overview of the findings from the focus group discussions can be found in Appendix 2.

Theoretical review

The purpose of the review of theory was to examine the theoretical body of knowledge that exists in relation to primary science learning, to examine what theories exist and how they help inform or interpret approaches to primary science education. The theoretical review thus fed into and complemented the systematic review.

The questions addressed by the theoretical review were:

- What theoretical models of learning exist of value to primary science teaching that could be useful to primary teachers of science?
- What are the mechanisms by which effective approaches work?

The theoretical review identifies the range of theories of relevance to primary science (e.g. behavioural, cognitivist, constructivist) with a view to establishing the extent to which they were reflected in interventions reported in the studies included in the systematic review.

'Effective' was taken to encompass the acquisition of knowledge; understanding; skills; and positive attitudes towards science/school science. However, the literature does not distinguish consistently between these, and very few studies consider all of them. It is also the case that much literature does not distinguish between pupils, except in terms of gender, so that it is difficult to say a great deal about disadvantaged learners.

The theoretical review takes the form of a narrative review and references were obtained from a range of sources:

- Existing knowledge of the primary science education literature within the team, including well-regarded sources of advice to teachers of primary science; and
- Additional literature obtained as a result of searches resulting from our existing knowledge, our review of practice, emerging findings from our systematic review, and suggestions from others.

Some of the theoretical models of learning of relevance to primary science teaching begin with the early years, while some are similar to theoretical models for other subjects or later phases of education, for example, the value placed on metacognition, cognitive load theory and self-regulation and on appropriate teacher feedback. The theoretical review has produced a critical overview of the rather large number of theoretical models (more modestly, these might be referred to as 'presumptions') of learning of particular relevance to primary science teaching.

The theoretical review was updated at two points. Firstly, it was updated at the conclusion of the research on practice, in order to ensure that it addressed theory that teachers reported as important in relation to achievement of outcomes. Secondly, it was updated towards the end of the systematic review to reflect studies including a mechanism for action based on theory.

The report on the theoretical review is found in Appendix 3.

The evidence map and systematic review

The evidence map formed the first phase of the systematic review and characterised all of the research on the topic from studies that met the specified inclusion criteria. The second phase of the systematic review involved narrowing the focus in order to ensure the review could answer the research questions with reference to those studies using approaches that could be replicated most readily in primary science in England, then assessing the included papers for their quality and synthesising the evidence.

The evidence map provided an overview of research into the teaching of primary science and informed decisions about which outcomes (e.g. attainment, attitudes) and approach type (e.g. curriculum, pedagogy, pupil, teacher and context) to take forwards to synthesis and analysis in the systematic review.

The question addressed by the evidence map was:

- What are the characteristics of existing evidence that point to the impact on pupil-level outcomes of different approaches to teaching primary science?

The evidence map presented approaches found in the literature, mapped against the outcomes for primary science pupils. The map told us what approaches and outcomes are referenced in the 169 studies included, and also where the evidence gaps are. The map was used to refine the review to ensure it was manageable, and to ensure the studies included were those most relevant to primary science in England. The research questions guiding the systematic review were:

- What are the most effective approaches for improving outcomes in primary science?
- What moderating factors influence the effectiveness of approaches?
- What are the barriers and facilitators to the effectiveness of approaches?
- What are the differential impacts of the approaches on identified disadvantaged pupils?

The methods used in the systematic review are described in the following section.

Report structure

In what follows, we describe the methods used in the systematic review. Thereafter, findings of the systematic review are discussed. These are organised first by cluster alphabetically (assessment and feedback; context-based and cross-curricular approaches; co-operative and collaborative approaches; critical thinking and argumentation; explicit instruction, mastery and modelling; ICT-supported and online teaching and learning; language, literacy and text-based approaches, learning outside the classroom, and practical work, inquiry and investigation. Within each, we describe and assess the evidence base. We then synthesise the findings across the clusters to identify the key conclusions and implications for policy and practice, along with an identification of gaps in the evidence base and limitations of the study.

A4 Methods

Introduction

The methods for the review of practice and the theoretical review have been summarised in the previous section of the report and can be found in detail in Appendices 1-3.

The following section of the report details the methods for the evidence map and systematic review.

Search strategy and results

The search strategy involved developing appropriate search strings to identify relevant research studies to answer the research questions. Studies were identified by (1) searching electronic bibliographic databases and (2) asking the EEF Guidance Panel to identify any documents relevant to the review. The following electronic databases were searched as they include studies focused on science education, primary education and psychology research literature: ERIC, ProQuest Dissertations and Theses, PsycInfo, SSCI and BEI. Google Scholar and JSTOR were not included in this process because of their limitations on the number of characters in search terms (150 and 200 respectively).

Account was also taken of differences in terminology in international literature, such as use of the term 'elementary', as well as 'primary', 'attainment' and 'assessment', 'attitude' and 'engagement'. The search terms included aspects mentioned by teachers in the focus groups undertaken as part of the Review of practice, and suggestions from the Guidance Panel.

Preliminary searches were undertaken in order to test the efficacy of the search strings. A large number of studies (thousands) were returned. Search terms were refined where necessary (e.g. adding a fifth string to avoid confounding terms).

The search strings were submitted to the EEF and the Guidance Panel for comment and the final list of search strings is presented in table 1. Wildcards are represented with an asterisk, and grouped words are enclosed in single quotes.

The full search term is:

String 1 AND string 2 AND string 3 AND string 4 NOT string 5.

Searches were conducted between 4th January 2022 and 2nd February 2022.

Table 2 shows the number of items retrieved from each of the bibliographic databases and manually added (e.g. identified from 'grey' literature or not found in the databases).

All records of full-text research studies identified in searches were uploaded. A total of 10458 unique sources were found and screened (see Appendix 4 for PRISMA diagram).

Table 1: Search strings

String	Search term
1	primary OR middle OR elementary OR 'early childhood' OR 'middle childhood' OR kindergar* OR 'key stage 1' OR 'key stage 2' OR 'Year 1' OR 'Year 2' OR 'Year 3' OR 'Year 4' OR 'Year 5' OR 'Year 6' OR 'grade 1' OR 'grade 2' OR 'grade 3' OR 'grade 4' OR 'grade 5' OR 'first grade*' OR 'second grade*' OR 'third grade*' OR 'fourth grade*' OR 'fifth grade*'
2	scien* OR biolog* OR chemi* OR physics OR 'physical science*'
3	learning OR curiosity OR attain* OR assess* OR outcome OR test OR grad* OR prog* OR feedback OR understand* OR knowledge OR enjoy* OR interest* OR relevan* OR excite* OR misconception OR preconception OR 'alternative understand*' OR attitud* OR engage* OR confidence OR disposition OR 'self-worth' OR career*
4	curricul* OR 'working scientifically' OR evidence OR classif* OR identif* OR 'pattern spotting' OR teach* OR pedagog* OR practic* OR investigat* OR skill OR inquir* OR enquir* OR question* OR collaborat* OR group OR discussion OR talk OR dialog* OR thinking* OR reason* OR vocabulary OR language OR linguistic OR metacogniti* OR 'self-regulat*' OR memor* OR 'cognitive strateg*' OR instruct* OR mastery OR argument*
5	'social scien*' OR 'physical education' OR 'primary care' OR 'medical scien*'

Table 2: items retrieved from bibliographic databases and identified from grey literature

Source	Total
BEI	1009
EEF database	40
ERIC	6553
Manually added items	12
ProQuest Dissertations	116
ProQuest documents	3302
PsycInfo	107
SSCI	4337
Total	**15476**

Inclusion and exclusion criteria for the review

In tandem with the development of the search strings, inclusion and exclusion criteria were developed for the systematic review to enable decisions to be made as to whether studies yielded in the searches should be included in the review. The review focused on 'approaches' to teaching primary science, so the inclusion/exclusion criteria covered both science-specific approaches (e.g., practical work, demonstration) or more general approaches to teaching and learning used in a science education context and relating to curriculum and/or pedagogy (e.g. metacognitive strategies, feedback). Extra-mural activities (e.g., summer schools and field trips) were also included if they related to science-specific contexts and/or outcomes.

The inclusion/exclusion criteria were submitted to the EEF Guidance Panel for comment. The final list of inclusion/exclusion criteria are found in Table 3.

Screening process

EPPI-Reviewer Web (beta) was used to store, organise and screen studies, and to facilitate production of the evidence map (Review ID 31129 EEF Primary Science). A coding tool was created in EPPI for screening:

- EXCLUDE on date
- EXCLUDE on age of population
- EXCLUDE on setting
- EXCLUDE on subject
- EXCLUDE on pupil outcomes
- EXCLUDE on counterfactual
- EXCLUDE as no intervention
- INCLUDE for second opinion
- INCLUDE on title and abstract.

For each study, the screener selected one option from the above.

A pre-screening phase involved members of the review team screening the same 31 studies on title and abstract. Discrepancies and interpretations to be taken forward to the screening phase were identified and discussed. An inter-screener reliability of 94% was calculated during this phase. The review team met first in pairs, then as a whole team to discuss disagreements and interpretations of the criteria, until there was 100% agreement on the 31 studies.

The first stage of screening was on the basis of title and abstract. All studies were distributed amongst the review team. Any studies coded 'INCLUDE for second opinion' were discussed with one lead screener. All studies marked for inclusion were also screened by a second reviewer, with any disagreements returned for screening and discussed between the first and second screener. This approach allowed decisions to be made efficiently and for less clear-cut decisions to be discussed with at least one other reviewer.

The second stage of screening involved retrieving studies for screening on full text. Studies were first distributed amongst the review team for screening on full text. All studies marked 'INCLUDE for second opinion' or marked for inclusion were double screened. The same codes as for screening on title and abstract were used, with the addition of INCLUDE on full text. The PRISMA diagram showing exclusions at each stage is presented in Appendix 4.

Table 3 Inclusion and exclusion criteria classified using PICOS and SPIDER (Methley et al., 2014).

PICOS	SPIDER	Scope	Explanation
Population	Sample	Children of primary age (5-11).	Studies that had a sample of children aged 5 to 11. Where studies involving children younger or older than 5-11 were included, only those samples where more than half were in the inclusion range were included in the review.
Intervention	Phenomenon of interest	Educational approaches to curriculum and pedagogy. Studies take place in educational settings.	Approaches included could be science specific (e.g., practical work, demonstration) or more general approaches to teaching and learning used in a science education context (e.g., metacognitive strategies, feedback), and may relate to curriculum or pedagogy. Summer schools and field trips were included where they relate to science-specific contexts and/or outcomes. The review focused on identifying approaches which improve outcomes in school settings, so to ensure ecological validity, findings must be relatable to these contexts. As a result, laboratory studies and theoretical studies were excluded.
Comparison	Design	Studies should include a counterfactual.	Qualitative and quantitative studies were included where they included a counterfactual. Process evaluations of randomised control trials (RCTs) which primarily consisted of qualitative and/or survey data were included where they meet the inclusion criteria above, or where the RCT they refer to meets the inclusion criteria.
Outcome(s)	Evaluation	All pupil outcomes identified from the review of practice are of interest.	Studies reporting pupil attainment, attitude and the other outcomes of interest identified in addendum 1 were included (e.g., attainment, attitude, conceptual understanding, confidence, interest, scientific language, thinking and working scientifically). All other outcomes were excluded.
Study design	Research type	Studies should include a counterfactual.	Qualitative and quantitative studies were included where they included a counterfactual.
Other criteria	Other	Published 2007-2021 Published in English	Studies that were published since 2007 and in English were included in the review.

Criteria for including studies in the evidence map were circulated to the EEF and Guidance Panel. Studies were included in the evidence map where:

- The study was published and included data collected January 2007 and September 2021.
- The study sample included pupils aged 5-11, or their teachers.
- The study was an approach to teaching, learning or assessing primary science.
- The study focused on the teaching of science.
- The study reported pupil outcomes.
- The study included a counterfactual.

In order to create an evidence map highlighting only key features of studies (principally, the approaches used and the outcomes reported). A 'light touch' data extraction tool was used to construct the map and shared with the EEF and guidance panel. The following codes were used:

- Approach to teaching primary science
- Outcomes for primary science pupils
- Type of study
- Publication type
- Country in which study was conducted

Method for moving from the evidence map to the systematic review

Each approach 'cluster' identified on the evidence map was reviewed during light touch data extraction. In consultation with the guidance panel, it was agreed that the review should include all approaches identified following screening. In order to refine the review to the agreed scope (100 studies at the upper end), the following refinements to existing criteria were made:

- Language: The study should focus on the teaching and learning of science (rather than language) and the outcome should report specifically on science. If not, exclude.
- ICT: The ICT application referenced in the study should be available in English, and accessible to teachers in England, or replicable in an English context with some effort. If not, or there is insufficient detail to replicate (for example, the study does not include details about how the ICT application is used in classrooms), exclude.
- Continuing Professional Development (CPD): This relates to any training, workshops, meetings, support or other approaches to capacity building or knowledge and skills development of teachers. CPD should be accessible to teachers in England, and have reported outcomes for primary aged children. Where it cannot be accessed, there is insufficient detail to replicate, or no pupil outcomes are reported, exclude.

Systematic reviews and meta-analyses were excluded in the search process. It was anticipated that the relevant primary studies in the reviews that met the inclusion criteria would have been identified by the search strings during the search.

In order to ensure exclusions based on the refined criteria were made consistently across the team, the following process was followed:

- Each included study was screened on full text using the refined criteria by one member of the review team.
- All studies marked for exclusion were screened on full text by a second member of the review team and (if different) the reviewer who made the decision on full text screening.
- Following discussion between the two/three reviewers, the decision was made, and the study retained or excluded.

At this stage, 90 studies remained in the review (Appendix 4).

Analysis approach

Clusters were determined by the review team in consultation with the EEF and the Guidance Panel, then revised and used to form theme headings for the report. These are:

1. Assessment and feedback
2. context-based and cross-curricular approaches
3. co-operative and collaborative approaches
4. critical thinking and argumentation
5. explicit instruction and related approaches
6. ICT-supported and online teaching and learning
7. language, literacy and text-based approaches
8. learning outside the classroom
9. practical work, inquiry and investigation

Data were extracted for analysis using the data extraction tool (Appendix 5). In the report, key features of studies are presented in tables, detailing contextual information (such as location and sample details), a description of the approach, measures used and reported pupil outcomes. Tables summarising studies are found in Appendix 7.

For each study, where possible, effect sizes are presented in the summary tables. Effect sizes are included as either reported in the study or calculated from data included in the study results. Where included or calculated, they are reported in the summary tables for each cluster. Where no effect size is reported or calculated, there was insufficient information to do so.

In tables presenting a summary of effects, the following key should be used:

- ES(R): Where no statistic stated, the effect size has been calculated by the authors of the study with no named statistic provided
- ES(C): calculated effect size
- *d*: Cohen's d
- η^2: Eta squared
- *g*: Hedges' g
- R^2: Square of Pearson correlation
- ES(R) *OR* = Odds ratio

It is appropriate to conduct a meta-analysis when studies in a cluster are sufficiently homogeneous in terms of participants, interventions, and outcomes, but if this is not the case, the meta-analysis is not very meaningful (Haidich, 2010). No meta-analyses have been conducted owing to the diversity of interventions, target outcomes and outcome measures in different subject areas within each cluster. Additionally, it was often possible to classify a study in a number of clusters, for example, where talk, collaboration, argumentation and inquiry were all features of approach. In these cases, the study was allocated to a single cluster based on 'best fit' following discussions within the review team. Where studies were found to be reporting the same dataset, they were placed in the same cluster for consistency.

Rather, each cluster includes a narrative synthesis to draw conclusions based on the evidence found within each cluster. This is organised according to the outcomes reported, so that all attainment outcomes within a cluster are discussed together. The synthesis includes a description of the evidence base (age of children, location, topics, approaches and outcome measures), identification of patterns of outcomes within each cluster, moderating and mediating factors, an assessment of the strength of the evidence supporting assessments, and a summary of findings for the approach.

Quality assessment

Each study was subjected to a risk of bias assessment. The quality of studies was assessed using an adaptation of the Effective Public Health Practice Project (EPHPP, n.d.) Quality Assessment Tool for Quantitative Studies (see Appendix 6) because it uses a scale that evaluates a range of study designs including RCTs and quasi-experimental studies. Although not designed for assessing the quality of educational studies, it includes assessments of study characteristics of relevance to this review on approach effectiveness. This allowed for the application of consistent criteria for assessing quality across a number of reviewers. All studies, regardless of quality assessment, are included in the review.

The quality assessment tool was based on the following: selection bias (likely representativeness of sample, proportion of those invited who participated in the study), study design (including method of randomisation and appropriateness of this method), differences between groups prior to intervention, reporting of withdrawals, control of confounders, blinding (whether or not the outcome assessors were aware of which pupils were in the treatment and control groups), whether pupils or teachers were aware of the research question, and reliability and validity of data collection tools. The summary assessments can be found in Appendix 8.

Part B: Systematic review findings

Systematic review findings: overview

Part B of the report presents the evidence from the systematic review. This systematic review is based on evidence from 90 studies.

Studies included in the review were organised into nine clusters of approaches (B1 - B9 below, in alphabetical order):

B1 Assessment and feedback
B2 Context-based and cross-curricular approaches
B3 Co-operative and collaborative approaches and peer teaching
B4 Critical thinking and argumentation
B5 Explicit instruction and related approaches
B6 ICT-supported and online teaching and learning
B7 Language, literacy and text-based approaches
B8 Learning outside the classroom
B9 Practical work, inquiry and investigation

The findings in each section include the following:

- A table containing entries for each study providing information about the population and context, description of the approach and reported outcomes. More detailed summaries of studies can be found in appendix 7.
- A description of the evidence base including the number, age of children, location, scientific topic, and description of approaches.
- A summary of the findings within the cluster or sub-cluster, with approach, control group, quality assessment, measures and effects on different outcomes reported or calculated. Note: care must be taken with interpretations of effect sizes. Few standard measures were used, and many used assessment instruments closely aligned to the content and nature of the intervention.
- Resourcing and implementation considerations and applicability of the evidence base to primary science teaching in England.
- Gaps in the evidence base.

B1 Assessment and feedback

Definitions

Assessment can be understood as the process of gathering and interpreting evidence in order to make a judgement about pupil outcomes (Harlen, 2007). Summative assessment is described as reporting or summarising attainment and can be a snapshot in time or a summary that takes a range of information into account (Earle, 2019). Formative assessment refers to the process of obtaining information from students, whether by teachers or the students themselves (Black *et al.* 2004), with the purpose of promoting learning (Black and Wiliam 2003). Integral to formative assessment is feedback, which Hattie and Timperley (2007, p.81) define as 'information that is provided from a range of different sources ... that relates to aspects of the learner's performance or understanding'. It has been observed that assessment can be difficult in science owing to the complexity of learning goals: whilst assessing vocabulary and calculation might be straightforward, assessing how conceptual ideas are applied during inquiry can be more difficult (Grangeat *et al*., 2021). Assessment methods used in primary science include tests for conceptual understanding, tracking grids, e.g. for inquiry skills, and other tasks, at times referenced against criteria (Earle & Turner, 2020). Feedback methods can include written or oral comments on work that result in action by the student (Wiliam, 2011).

Description of the evidence base

Table 5 summarises the key features of each study. A more detailed summary is found in appendix 7. Six studies explored assessment and feedback, of which five focused on the use of formative assessment and one on summative assessment.

A variety of approaches to assessment are evaluated in the studies. These included the use of audible image description, embedded formative assessment and teacher feedback, scaffolding instructional discourse, peer-assisted learning, guided peer feedback and bidirectional peer assessment using concept maps. Embedded formative assessment refers to the planned use of structured evaluations of students' progress, implemented at specific moments at which an important learning goral should be met (Decristan *et al.,* 2015). Scaffolding instructional discourse relates to talk where the teacher supports the pupil to solve problems or complete tasks that they would otherwise be unable to achieve (Decristan *et al.,* 2015). In guided peer feedback, pupils play the role of reviewer based on criteria created by the teacher (Hwang et al., 2018). This is bidirectional when pupils both give and receive reviews of their work (Hwang et al., 2021).

The interventions involved summative and formative assessment by teachers and by peers, with pupils aged 8 to 13 years. The studies were conducted in a broad range of science topic areas including buoyancy, geology, plants and ecology. No studies were conducted in the UK, three were undertaken in Germany[2], two in Taiwan[3] and one in the USA.

Three studies adopted a quasi-experimental design and three were randomised control trials. Through a quality assessment tool (see appendix 6), five studies were rated as moderate and one as low-quality. In all cases but one, the control condition was 'business as usual' without the formative assessment intervention. In Hwang *et al.* (2018), the control was teacher feedback (rather than peer feedback).

---

[2] These three studies have researchers in common. Decristan *et al*. (2015a) and Hondrich *et al*. (2018) appear to be based on the same sample.

[3] These two studies have the same lead author.

Five studies focused on attainment (knowledge and conceptual understanding) as an outcome, with two reporting data on attitudes, interest and motivation and one on critical thinking. A range of measures were used to measure attainment. Studies by Decristan *et al.* (2015a;b) used a competence test based on TIMSS 2007 and those by Hwang *et al.* (2018) and Hwang *et al.* (2021) used a task-specific assessment rubric designed by the research team and a test developed by teachers respectively.

Findings

Most studies reported positive outcomes for pupils participating in interventions on assessment compared with pupils who did not participate in the intervention. Few of the studies reported effect sizes. These were calculated from the data where possible.

Attainment (knowledge and conceptual understanding)

The single study focusing on summative assessment (Ferrell *et al.*, 2017) found improved outcomes for students who were given an audible image description during summative assessment for pupils who read braille, with no effect on pupils with print disabilities or pupils with visual disabilities who read print. Whilst there were no gains in attainment for those who can read print with difficulty, there were no negative impacts reported.

Four studies involving approaches to formative assessment reported a positive impact of interventions (embedded formative assessment, scaffolded instructional discourse, guided peer feedback and bidirectional peer feedback) on attainment. Decristan *et al.* (2015b) found a positive impact reported for formative assessment with no impacts reported for peer-assisted learning or structured instructional discourse.

Moderating effects identified included language proficiency (Decristan *et al.* 2015b). Scaffolded instructional discourse and formative assessment was found to be particularly beneficial for students with low language proficiency in the language of instruction (Decristan *et al*., 2015b). Other moderating effects reported by studies in this cluster included a supportive classroom environment, good classroom management, and cognitive activation (Decristan *et al*., 2015a).

Attitudes, interest and motivation

Two of the six studies collected data on attitudes, motivation, and interest (Hondrich *et al*., 2017; Hwang *et al*., 2021), both reporting positive effects of the formative assessment intervention.

The study conducted by Hondrich *et al*. (2017) measured the impact of providing teachers with an adapted unit of teaching focused on floating and sinking with embedded formative assessment material (a written task, semi-standardised individual feedback and adapted worksheets). For unit two, teachers had to develop the formative assessment themselves. The study found a small positive effect on intrinsic motivation of students after both units.

Hwang and Chang (2021) investigated the effect of pupils responding to peer feedback using concept maps on self-efficacy and learning motivation, measured using questionnaires adapted from existing instruments. A small positive effect was found on self-efficacy, with no difference in learning motivation between the control and experimental groups.

Perceived competence was found to moderate pupils' intrinsic motivation (Hondrich *et al*., 2018).

Table 5: Key features of studies on assessment and feedback in primary science teaching

Authors	**Country**	**Design**	**Age of pupils**	**Sample** pupils *schools*	**Approach and assessment**	**Comparison**	**Outcomes measured**	**Summary of effects**
Decristan *et al.* (2015a)	Germany	Cluster RCT	8-9	551 *18*	Embedded formative assessment (floating and sinking) QA result: moderate	Teaching as usual: inquiry unit (floating and sinking)	Attainment: Science competence test based on TIMSS 2007	*Attainment* ES(R): R^2=0.12
Decristan *et al*. (2015b)	Germany	Cluster RCT	8-9	873 *39*	SID: scaffolding instructional discourse FA: formative assessment PAL: peer assisted learning All in floating and sinking QA result: moderate	Teaching as usual: inquiry unit (floating and sinking)	Attainment: Science competence test based on TIMSS 2007	*Attainment* Data presented do not facilitate calculation of ES.
Ferrell *et al.* (2017)	USA	Quasi-experiment	8-13	295 *not stated*	Summative assessment: audible image description Various topics QA result: low	No audible image description (various topics)	Attainment: Test based on items selected from the Utah Test Item Pool Server	*Attainment* ES(R)=0.66 (pupils who cannot read print) Pupils who can read print with difficulty: no effect
Hondrich *et al*. (2018)	Germany	Cluster RCT	8-9	551 *18*	Embedded formative assessment (floating and sinking) QA result: moderate	Teaching as usual: inquiry unit (floating and sinking)	Attitudes: Questionnaires assessed by scales adapted from Blumberg (2008)	*Attitudes* Perceived competence ES(C): *g*=0.24 Intrinsic motivation ES(C): *g*=0.37
Hwang *et al*. (2018)	Taiwan	Quasi-experiment	10	72 *1*	Guided peer feedback during an e-book development activity (plants and ecology) QA result: moderate	Teacher feedback during an e-book development activity (plants and ecology)	Attainment: E-book assessment rubric developed by researchers and teachers Critical thinking: Innovative thinking tendency scale modified from Lin and Wang (1994)	*Attainment* ES(C): *g*=5.42 *Critical thinking* ES(C): *g*=0.52
Hwang & Chang (2021)	Taiwan	Quasi-experiment	10	101 *1*	Bidirectional peer assessment (geology) QA result: moderate	Peer assessment with no opportunity to respond (geology)	Attainment: Test developed by experienced science teachers Attitudes: Questionnaires on Learning motivation (Wang & Chen, 2010) and self-efficacy adapted from Pintrich, Smith, Garcia, and McKeachie (1991) Critical thinking: Critical thinking tendency questionnaire adapted from Chai, Deng, Tsai, Koh, and Tsai (2015)	*Attainment* ES(C): *g*=0.51 *Attitudes* Learning motivation no effect *Critical thinking* ES(C): *g*=0.92

Critical thinking and argumentation

Two out of the six studies collected data on outcomes relating to critical thinking, both using peer feedback in online contexts. Questionnaires were used to measure innovative thinking (Hwang *et al.* 2018), cognitive load (Hwang *et al.* 2018; Hwang &Chang, 2021) and critical thinking tendency (Hwang & Chang, 2021). These peer feedback approaches were associated with reduced cognitive load and medium to large positive effects on innovative thinking and critical thinking tendency.

The findings support the generally held view that formative assessment allows for a greater number of interactions, and space to think, discuss and consolidate learning for pupils.

Implementation and resources

Three studies (Decristan *et al.*, 2015a; Decristan *et al.*, 2015b; Hondrich *et al.*, 2018) included professional development for teachers. None of the studies reported costs, but there is likely to be a cost associated with training for teachers and for some of the online tools (ebooks and concept mapping software) used in the studies by Hwang *et al.* (2018) and Hwangand Chang (2021). These latter studies overlap with ICT-based and online approaches, although feedback was the central focus of these studies.

Gaps in the evidence base

Overall, the evidence base is very small, with five studies focusing on formative assessment, and one on summative assessment. The practice review found teachers reported feedback and assessment to be an area in which they lacked confidence. Given that there is currently a statutory requirement for teachers to make assessment judgements on science at the end of Key Stage 2, there is a need for further evidence in this area.

Additionally, the outcomes of focus are narrow (mainly attainment, with some capturing motivation and critical thinking). No studies reported on the impact of assessment on outcomes relating to scientific language and communication, or to working scientifically and problem solving. Other gaps in the evidence base include studies involving younger children (of Key Stage 1 age), studies conducted in England during the review period, and differential outcomes of interventions for pupils. There is some tentative evidence to indicate particular benefits of formative assessment for pupils with lower language proficiency, but this would need further exploration to substantiate.

Quality of the evidence base

Of the six studies included in this cluster, five were assessed as moderate-quality and one low-quality.

All interventions were applied at the classroom level and in schools. The three studies conducted in Germany have higher numbers of participants and were conducted in >10 schools, in contrast to the two studies conducted in Taiwan (both in a single school), and one in the USA, where the number of schools is not stated.

A principal feature of this cluster is the diversity of focus, interventions and measures used to assess effectiveness, unless studies have been carried out by the same research team. This applies to both achievement and attitudes. This absence of homogeneity meant that meta-analysis was not possible.

Applicability of the evidence base

Assessment was identified as a priority by teachers in England who participated in the review of practice (Appendix 2), in particular the assessment of practical skills. Teachers reported using teacher-, peer- and self assessment approaches. Whilst the approaches reported by Decristan *et al*. (2015a & 2015b), Hondrich *et al*. (2018) and Hwang *et al*. (2018 & 2021) do not appear to be country-specific, the review of practice suggests that there is widespread formative assessment practice in England and it may be sensible to prioritise research on the effectiveness of approaches already in use. Formative assessment has been integral to a number of national programmes, notably the Primary Science Quality Mark (PSQM) and Teacher Assessment in Primary Science (TAPS) project – published after the timeframe for this review – suggesting that evidence on assessment is likely to be relevant to teachers in England.

Overall evidence statement/Key findings

Evidence suggests that formative assessment can have a positive effect on attainment and critical thinking, and a null to positive effect on attitudinal outcomes such as motivation, perceived competence and self-efficacy. This supports evidence from systematic reviews which have found positive effects of formative assessment on student learning in other areas (e.g. Lee *et al*., 2020). A number of the studies would appear to have the potential to be adapted to the English context and subjected to trials to determine 'proof of principle'. Table 6 provides a summary of the evidence.

Table 6: Summary of the evidence on assessment and feedback

Number of studies	6
Assessed quality of studies	High: 0 Moderate: 5 Low: 1
Location of studies	Germany: 3 Taiwan: 2 USA: 1
Outcomes of focus	Attainment (knowledge and conceptual understanding): 5 Attitudes, motivation and interest: 2 Critical thinking and argumentation: 2
Design	Quasi-experimental design: 3 Randomised control trial: 3
Consistency of findings	Results relating to the impact of formative assessment on attainment were consistently positive when compared with business-as-usual. For motivation, one study found a positive impact while another reported no impact (Hwang *et al.* 2021)
Summary of the evidence base	• Small number of studies • Study design: three studies used a randomised control trial design • Three large-scale moderate-quality studies • Variability in interventions and outcome measures. • All studies except one developed their own instruments, rather than using standardised measures • All interventions were developed and evaluated by the researchers reporting the work

B2 Context based and cross-curricular approaches

Definitions

Context-based approaches are those in which scientific concepts and process skills are applied in real life contexts relevant to pupils from diverse backgrounds (Kuhn & Müller, 2014). Context-based approaches include approaches where concepts are taught through a context, where a narrative is used to frame scientific discoveries or where pupils and teachers inquire into a topic important to their community (Gilbert, 2011). Examples of context-based approaches include the use of newspaper stories as contexts for learning about specific scientific ideas and the use of school gardens or greenhouses to learn about plants. Such an approach aims to make pupils' experiences in the science lessons more appealing and relevant (Ramsden, 1997). *Cross-curricular approaches* draw on more than one subject or discipline with a view to helping pupils see the links between the areas and make their learning more relevant and engaging.

Description of the evidence base

Table 7 summarises the key features of each study. Six studies focused on teaching and interacting science using real world contexts (Burt *et al.*, 2022; Olgun & Adali, 2008; Zhang & Campbell, 2012; Fasasi, 2017) or using cross-curricular/interdisciplinary approaches (Hardiman *et al.*, 2017; Qiao & Zhou, 2020). The duration of interventions varied from hours (Qiao & Zhou, 2020), to several weeks (Fasasi, 2017; Hardiman *et al.*, 2017; Olgun & Adali, 2008) to one year (Burt *et al.*, 2022; Zhang & Campbell, 2012).

Context-based approaches included ethnoscience instruction in which the prior cultural beliefs and ideas of learners are accessed and related to the scientific concept being taught (Fasasi, 2017), the use of hydroponic gardening to teach a climate change curriculum (Burt *et al.*, 2022), the use of case studies to teach about bacteria, fungi and protista (Olgun & Adali, 2008) and an integrated experiential learning curriculum in which pupils investigate a question, problem or situation relevant to them (Zhang & Campbell, 2012). Cross-curricular approaches included arts-integrated instruction (Hardiman *et al.*, 2017) and STEM teaching (Qiao & Zhou, 2020).

There was little consistency in underpinning theories of change. Theoretical approaches explicit in reports included constructivism (Olgun & Adali, 2008) and theories of memory (Hardiman *et al.*, 2019). Hardiman *et al.* (2019) draw on cognitive science as a basis for their arts-integrated approach, arguing that activities inherent in the arts (e.g. repetition, elaboration, emotional arousal and representing through images) support the improvement of memory in science.

The studies were conducted in a range of science topics, with examples from biology, chemistry and physics. Studies were conducted in China, Nigeria, Turkey and the USA, with pupils aged between 8 and 12 years old.

All studies adopted a quasi-experimental design except one, which was a randomised control trial (Hardiman *et al.*, 2017). Five studies were assessed to be moderate-quality and one was assessed to be low-quality.

Three of the studies focused on attainment (Qiao & Zhou, 2020; Burt *et al.*, 2022; Hardiman *et al.*, 2017). Two focused on attitudes (Zhang & Campbell, 2012; Fasasi, 2017) and one explored both dimensions (Olgun & Adali, 2008). In terms of attainment, Burt *et al.* (2022) measured impact using an existing state-wide test, whereas Qiao and Zhou (2020) and Olgun and Adali (2008) developed their own tests.

In terms of attitudinal measures, Fasasi (2017) adapted an existing scale (Deepken, Lawsky & Padwa, 2003), Olgun and Adali (2008) used a published scale (Sahin, Çakır, & Sahin, 2000) and Zhang and Campbell (2012) used their own scale (Zhang & Campbell, 2010).

Findings

Most studies reported statistically significantly higher attainment and/or attitudes for pupils participating in their interventions compared with pupils who did not participate in the intervention.

Care must be taken with interpretation of effect sizes as it was not always clear what the control group did (e.g. Burt *et al.*, 2022). Where the control group was given 'business as usual' (Zhang & Campbell, 2012; Qiao & Zhou, 2020), this was not often described. Hardiman *et al.* (2017) reported the use of a conventional science instruction unit in their comparison with an arts-integrated instructional unit. Olgun and Adali (2008) reported that the control group did a reading assignment with teacher question and answer, and Fasasi (2017) reported a modified lecture method with teacher demonstrations used with the control group.

Attainment (knowledge and conceptual understanding)

Four studies reported outcomes on attainment or conceptual understanding for context-based and cross-curricular approaches, although there was considerable diversity in approach and outcome. All reported positive effects, except the arts-integrated instruction (Hardiman *et al*., 2017), which found no difference between groups, indicating that whilst there was no gain in content retained, nor were students experiencing arts-integrated instruction disadvantaged.

Three studies were found to have positive effects. Burt *et al*. (2020) reported on a large-scale experimental study focused on a climate change curriculum including hydroponic gardening, worm-composting, use of rainwater catchment systems and pest management. They found that the experimental group performed better than the control group in state science tests. Olgun and Adali (2008) found gains in attainment for pupils in a group that used a case study approach to real-world problem solving in the context of viruses and bacteria compared to those who were given a reading assignment prior to lessons with time used for one child to explain key concepts and the pupils writing down the teacher's explanation. The STEM teaching intervention on buoyancy, described as interdisciplinary, technologically enhanced, project-based and hands-on, was reported to have a large positive effect when compared with 'traditional teaching' (Qiao & Zhou, 2020).

There was some evidence in the study on arts-integrated instruction (Hardiman *et al*., 2017) to suggest that reading level was a moderating effect, with basic readers retaining more in arts-integrated instruction. This study also found that order matters: pupils who had arts-integrated instruction first remembered more science later.

Table 7: Key features of studies on context-based and cross curricular approaches to primary science teaching

Authors	**Country**	**Design**	**Age of pupils**	**Sample** pupils *schools*	**Approach and assessment**	**Comparison**	**Outcomes measured**	**Summary of effects**
Burt *et al*. (2022)	USA	Quasi-experiment	9-10	3121 *28*	New York Sun Works (NYSW) Programme: project-based curriculum with hydroponic gardening (climate change and sustainability) QA result: moderate	Demographically matched and delayed treatment comparison group, not specified what they did.	Attainment: The New York State science achievement tests	*Attainment* ES(C): g=0.55 (demographically matched group) ES(C): g=0.64 (unmatched group)
Fasasi (2017)	Nigeria	Quasi-experiment	9-12	352 *4*	Ethnoscience instruction linking science education to culture. Includes sharing cultural beliefs, sayings and practices relevant to the topic, and classification of these as compatible, modifiable or contradictory to science (various topics) QA result: moderate	Modified lecture method, drawn from the basic science curriculum module of the Federal Ministry of Education	Attitudes: Attitude Towards Science Scale adapted from standardised Modified Sherman Science Attitude Scale (Doepken, Lawsky & Padwa, 2003)	*Attitudes* ES(R): η^2=0.46
Hardiman *et al*. (2017)	USA	RCT	10-11	350 *6*	Arts-integrated instruction involving demonstration of knowledge through visual and performing arts (various topics). QA result: moderate	Conventional science instruction unit matched for content, dosage and activity type	Attainment: Multiple choice content assessment designed by research team	*Attainment* No difference between groups reported. Data presented do not facilitate calculation of ES.
Olgun and Adali (2008)	Turkey	Quasi-experiment	9-10	88 *1*	Real-life problem solving using internet research and classroom discussion (viruses and bacteria) QA result: low	Reading assignment prior to lesson with one pupil reporting key constructs, then teacher explains and asks questions	Attainment: Multiple choice Science Achievement Test (SAT) designed by researchers Attitudes: Attitude Scale Towards Science (ASTS) developed by Sahin, Çakır, and Sahin (2000)	*Attainment* ES(C): g=1.32 *Attitudes* ES(C): g=1.02
Qiao and Zhou (2020)	China	Quasi-experiment	10-11	200 *1*	STEM education integrating science, technology, engineering and mathematics (buoyant force): QA result: low	'Traditional science teaching' (no further details)	Attainment: Basic knowledge and ability expansion questionnaire designed by researchers	*Attainment* ES(C): g=0.93
Zhang and Campbell (2012)	China	Quasi-experiment	8-11	385 *10*	Problem solving (Integrated Experiential Learning Curriculum, IELC) (topic not stated) QA result: moderate	'Traditional science lessons' non-IELC (no further details)	Attitudes: Teacher Attitude about Teaching Science instrument designed by researchers	*Attitudes* ES(R): R^2=0.18

Attitudes, interest and motivation

Three studies reported positive impacts of context-based and cross-curricular approaches on attitudinal outcomes. Problem-solving approaches (Olgun & Adali, 2008; Zhang & Campbell, 2012) had positive effects on attitudes compared with a pre-set reading task and teacher explanation and 'traditional science teaching' respectively. Fasasi (2017) found improved attitudes towards science when ethnoscience instruction was used compared with lectures combined with demonstrations. This approach encouraged pupils to make connections between science education and their culture and language, and was particularly positive for children in rural schools and those with parents with low educational status.

These studies suggest that curricula emphasising real world contexts and problem solving improve attainment and attitudes towards science and the environment.

Implementation and resources

None of the studies reported barriers to implementation or costs. However, two approaches (Burt *et al*., 2022; Zhang & Campbell, 2012) involved training for teachers. The climate change curriculum (Burt *et al*., 2022) required modification of classrooms, and application of this approach would require investment in hydroponic gardening equipment and training to be applied in schools in England.

Any potentially promising intervention would need to be adapted sensitively to the English context and tested to determine proof of principle.

Gaps in the evidence base

Few studies investigated the impact of context based and cross-curricular approaches, with none conducted in England. The review of practice found that cross-curricular and thematic curriculum planning and teaching sensitive context was in use, so studies on the effectiveness of these approaches are likely to be of interest to teachers.

Studies in this cluster focused on a narrow range of outcomes, namely attainment and conceptual knowledge and understanding, and attitudes. There was little attention to the effect of context-based and cross-curricular approaches on critical thinking and argumentation, scientific language and communication or working scientifically and problem solving. No studies focused on children younger than 8. Furthermore, there was little attention to differential outcomes for different groups of pupils in this cluster. Whilst there was some exploration of mediating factors such as reading skill (Hardiman *et al.*, 2017; Burt *et al.*, 2022), these were not drawn on in any substantive way in the discussion of the findings.

Quality of the evidence base

Five studies were assessed to be of moderate quality and one of low quality. All interventions were applied at the classroom level in schools. Several studies involved only one or two schools (Fasasi, 2017; Olgun & Adali, 2008; Qiao & Zhou, 2020). Although there is consistency in findings, there is a small number of studies with varying approaches to context and cross-curricular approaches, with only one RCT. A principal feature of this cluster is the diversity of focus (on science topic and intended outcomes), intervention, theoretical basis and measures used to assess effectiveness. This applies to both achievement and attitudes. To some degree, this is related to the diversity of interpretation of the term 'context'.

Applicability of the evidence base

No studies in this cluster were conducted in England, although teachers who participated in the review of practice reported using context-based and cross-curricular approaches.

Problem-solving, arts-integrated instruction, STEM teaching, ethnoscience instruction and a climate change curriculum could all be applied to teaching in England, but caution is needed because these approaches have been tested in countries with different education systems, curricula and approaches to pedagogy, with 'traditional teaching' often not described.

Overall evidence statement/Key findings

The evidence suggests that context-based and cross-curricular approaches can have a positive effect on pupil attainment and on attitudes. There is some tentative evidence to suggest that there are particular benefits for pupils with lower reading skills and for those with parents who have had less education. Table 8 provides a summary of the evidence.

Table 8: Summary of the evidence on context-based and cross-curricular approaches

Number of studies	6
Assessed quality of studies	High: 0 Moderate: 4 Low: 2
Location of studies	China: 2 Nigeria: 1 Turkey: 1 USA: 2
Outcomes of focus	Attainment (knowledge and conceptual understanding): 4 Attitudes, motivation and interest: 3
Design	Quasi-experimental design: 5 Randomised control trial: 1
Consistency of findings	Results were consistently positive when compared with business-as-usual with the exception of one dimension (attainment) for one approach (arts-integrated instruction (Hardiman *et al.*, 2017), which found no difference between experimental and control groups.
Summary of the evidence base	• Study design: one randomised control trial • Consistency in findings • Few large-scale high-quality studies • Variability in interventions and outcome measures • All studies except one developed their own instruments, rather than using standardised measures • All interventions were developed and evaluated by the researchers reporting the work

B3 Cooperative and collaborative approaches

Definitions

Collaborative and cooperative learning refers to classroom techniques in which groups of pupils learn or attempt to learn together jointly (in collaborative learning) or where the work is shared systematically (in cooperative learning) (Dillenbourg, 1999). Groups can be considered as those containing two or more children, although there is discussion about whether pairs do constitute a group (Schoor, Narciss & Körndle, 2015). During cooperative and collaborative learning, both self- and social-regulation can occur (Schoor, Narciss & Körndle, 2015).

Description of the evidence base

Six studies featured approaches that involve cooperative or collaborative approaches. Cooperative and collaborative approaches included collaboration during differentiation (Eysink *et al.*, 2017) and collaborative discussion, such as the self-explain-discuss-re-explain strategy, an approach to support pupils to think independently through discussion and explanation (Chang & Hsin, 2021) as well as the use of software-based sticky notes (Looi *et al.*, 2010). This cluster also included argumentation through collaborative writing using the science writing heuristic, a structured approach to writing which links questions, evidence and claims (Hand *et al.*, 2018; Chen *et al.*, 2013; Reeves *et al.*, 2013).

Studies in this cluster included children aged 7-11. The studies were undertaken in the Netherlands, Singapore, Taiwan (all one study) and the USA (three studies). Two studies (Chang & Hsin, 2021; Chen *et al*., 2013) were set in a single topic context (forces and motion), with the others used in multiple scientific topics.

Four studies adopted a quasi-experimental design and two were randomised control trials. Using a risk of bias tool, one study was assessed as high-quality, four as moderate and one as low-quality owing to how teachers opted into the intervention group and differences between the intervention and control group.

All six interventions in this cluster focused on attainment. Other outcomes reported included attitudes (Looi *et al.*, 2010), scientific language and communication (Chen *et al.*, 2013; Reeves *et al.*, 2013) and critical thinking and argumentation (Chen *et al.*, 2013; Hand *et al.*, 2018; Reeves *et al*., 2013).

Outcome measures included standardised tests (Hand *et al.,* 2018; Reeves *et al.*, 2013) of science and critical thinking, a researcher-developed attitude questionnaire (Looi *et al.*, 2010) and concept test (Eysink *et al.*, 2017), an externally designed topic test and argumentative assessment rubric designed by the researchers (Chen *et al.*, 2013).

Given the nature of approaches and the outcomes of focus in this cluster, there are strong links with the *language and literacy*, and the *critical thinking and argumentation* clusters (particularly with reference to the science writing heuristic).

Findings

Table 9 presents an overview of the effects of approaches in the cooperative or collaborative approaches cluster.

Table 9: Key features of studies on co-operative and collaborative approaches to primary science teaching

Authors	Country	Design	Age of pupils	Sample pupils *schools*	Approach and assessment	Comparison	Outcomes measured	Summary of effects
Chang & Hsin (2021)	Taiwan	Quasi-experiment	10-11	104 *1*	Worksheets followed by distinct phases of self-explain, discuss and re-explain (position of the sun) QA result: moderate	Same worksheets with pupils able to ask questions about what they didn't understand	Attainment: Two-tier multiple-choice test developed on the basis of the Teacher's Manuals for Elementary School Science and Technology	*Attainment* ES(R): *d*=0.71 for low-achievers ES(R): *d*=0.61 for high-achievers
Chen, Hand & McDowell (2013)	USA	Quasi-experiment	9-10	838 *4*	Writing-to-learn. Letter exchange between pupils with focus on argumentation (forces and motion) QA result: moderate	Standard curriculum (teaching as usual)	Attainment: Multiple-choice questions developed by Horizon Research Institution (developer of tests for use in schools)	*Attainment* ES(C): *g*= 0.25
Eysink, Hulsbeek & Gijlers (2017)	Netherlands	Quasi-experiment	8-11	306 *11*	STIP approach, translated as Collaboration during differentiation in Task, Content, and Process (various topics) QA result: low	Regular instructional approach with textbook and exercise book. Control group had more experienced teachers	Attainment: Researcher-developed tests used to measure pupils' domain knowledge	*Attainment* ES(R): *d*=0.46 in favour of control group
Hand *et al.* (2018)	USA	Cluster RCT	7-10	9963 *48*	Science Writing Heuristic (various topics) QA result: high	Standard curriculum (teaching as usual)	Attainment: Iowa Test of Basic Skills, Iowa Assessments Test Critical thinking: Cornell Critical Thinking (CCT) test	*Attainment* No educationally significant effect *Critical thinking* ES(C): *g*=0.17
Looi, Chen & Ng (2010)	Singapore	Quasi-experiment	10	240 *1*	Collaborative activities with Group Scribbles online sticky note (various topics) QA result: moderate	No participation in collaborative Group Scribbles	Attainment: Exam designed by teachers Attitudes: Researcher-developed questionnaire about attitudes towards science learning	*Attainment* ES(R): η^2=0.07 *Attitudes* Interest in group work ES(C):*g*=0.25 Decreased interest in working individually ES(C):*g*=0.45 Science learning ES(C):*g*=0.27
Reeves *et al.* (2013)	USA	Cluster RCT	9-10	4713 *48*	Science Writing Heuristic as part of an inquiry-based approach (various topics) QA result: moderate	Inquiry-based approach without science writing heuristic	Attainment: Iowa Test of Basic Skills science subgroup	*Attainment* No effect on science scores

Attainment (knowledge and conceptual understanding)

All six studies in this cluster reported on attainment in science, with mixed effects. These are discussed by approach: collaboration through discussion, argumentation through collaborative writing and collaboration through differentiation. Collaboration through discussion refers to approaches involving talk between pupils (e.g. Chang & Hsin, 2021). Argumentation through collaborative writing includes approaches where pupils discuss and write in order to persuade others (e.g. Chen *et al.*, 2013; Hand *et al.*, 2018). Collaboration through differentiation involves all pupils working together on the same theme with varying resources, tasks and levels of teacher support (e.g. Eysink *et al.*, 2017).

Two small-scale studies found a positive effect on attainment of collaboration through discussion. The Self-explain-Discuss-Re-explain strategy (Chang & Hsin, 2021) which gave pupils time to think and discuss ideas with others had a positive effect on pupils' conceptual understanding, with low achievers gaining more than high achievers. This closed the achievement gap whilst raising achievement overall. Similarly, the quasi-experiment by Looi *et al.* (2010) on collaborative activities through Group Scribbles provided some evidence that collaborative classroom talk can play a role in improving attainment in science.

The evidence is mixed when it comes to the impact on science attainment of argumentation-driven approaches. The science writing heuristic (Keys *et al*., 1999) is a model for teachers and students to negotiate meaning about investigative activities with a particular focus on argumentation (claims, evidence and reflection). It is based on social constructivist principles and involves phases of individual and group exploration, negotiation and writing. The large-scale randomised controlled trials undertaken in the USA (Hand *et al*., 2018; Reeves *et al*., 2013) found no significant gains and low treatment path coefficients for science attainment. It is worth noting that Hand *et al.* (2018) employed widely used national tests to assess the impact on attainment of their intervention, while Chen *et al.* (2013) used externally developed tests. In the writing-to-learn study (Chen *et al.*, 2013) in which pupils wrote to older peers, better performance was reported for girls, pupils eligible for free or reduced price school meals, pupils who have an individualised education programme and those who are on a gifted individualised education programme. However, the reported effects were trivial or small.

The single study reporting a negative impact of collaborative approaches (Eysink *et al.*, 2017) used an approach to collaboration during differentiation. This places pupils in mixed groups based on prior test results to work towards a common goal. In this quasi-experiment, teachers were able to opt into the experimental or control group. There was a marked difference in the average age of the teachers in the experimental group (23) compared to the control group (44). The researchers re-analysed the data focusing on teachers who had scored highly on the differentiation activity based on observations, i.e. where there was higher treatment fidelity, with a reversed result (i.e. the treatment group outperformed the control group).

Attitudes, interest and motivation

One study reported attitudinal outcomes. This study focused on the collaborative use of Group Scribbles online sticky notes (Looi *et al.*, 2010). The authors reported relatively small positive effects on pupils' interest in group work and in science learning, with decreased interest in working individually.

Critical thinking and argumentation

One study reported positive effects on critical thinking and argumentation: the science writing heuristic (Hand *et al.,* 2018), which is focused on teaching argumentation. This approach involved small group learning using a framework for argumentation. Working with a subset of participants, Hand *et al*. (2018), reported significant gains in critical thinking, particularly for pupils with individual education plans, those eligible for free or reduced school meals, and those who speak English as an additional language.

Implementation and resources

Implementation factors and barriers were not discussed in depth. Several studies involved CPD for teachers or required access to hardware and software, but costs were not reported.

The interventions of both Hand *et al.* (2018) and Chen *et al.* (2013) had a substantial teacher professional development component associated with them, and all were interventions that happened over an extended period of time. This suggests there is some work to be done in England to test 'proof of principle' with the use of resources aimed to develop primary-age pupils' skill in using argumentation and develop critical thinking abilities, together with the CPD needed to support teaching with such an intervention.

Other resources included online software (for Group Scribbles) which is likely to have a cost.

Gaps in the evidence base

The studies reported here tended to use cooperation and collaboration as a feature of a broader approach such as use of the science writing heuristic. No studies reported on outcomes related to working scientifically and problem solving, and only one study (Hand et al., 2017) focused on children of Key Stage 1 age, but those pupils were not included in the critical thinking testing. Given that there is considerable interest in co-operative and collaborative learning in primary science, further work into the nature of effective group work would be of benefit to the field.

Quality of the evidence base

This cluster consists of studies assessed high- (1), moderate- (4) and low-(1) quality. There are two large scale RCTs conducted in a relevant educational context. The study by Hand *et al.* (2018) was a large-scale RCT conducted in a large number of classrooms in the USA. The large RCT by Reeves *et al.* (2013) reported in a conference paper provides fewer details. The other studies were smaller in scale and tended to use researcher-designed instruments for their evaluation. The studies using cooperation and collaboration were also very diverse in nature.

Applicability of the evidence base

The approaches in this cluster are broadly applicable to England as they use worksheets and teaching strategies that could be readily implemented. The study by Looi *et al.* (2010), looking at the use of computerised sticky notes, draws on technology in science lessons. While individual technology programmes may not be directly transferable from one context to another, their use signals an interest in the area and points to the desirability of work being undertaken to assess the features of such programmes which help promote effective learning.

Overall evidence statement/Key findings

There is mixed evidence in the co-operative and collaborative approaches cluster. In terms of attainment, two small-scale studies found a positive effect of collaboration during discussion, with null or positive effects of collaboration during discussion, suggesting that these approaches at least appear to do no harm to science attainment. Positive effects were found on attitudinal outcomes, and on critical thinking. However, these were the result of one small-scale study and one larger-scale study respectively. Carefully structured interventions involving cooperation or collaboration appear to have particular benefits for disadvantaged pupils. Table 10 provides a summary of the evidence.

Table 10: Summary of the evidence on cooperative and collaborative learning approaches

Number of studies	6
Assessed quality of studies	High: 1 Moderate: 4 Low: 1
Location of studies	The Netherlands: 1 Singapore: 1 Taiwan: 1 USA: 3
Outcomes of focus	Attainment (knowledge and conceptual understanding): 6 Attitudes, motivation and interest: 1 Critical thinking and argumentation: 1
Design	Quasi-experimental design: 4 Randomised control trial: 2
Consistency of findings	Some mixed findings, with null (Hand *et al*., 2017) and positive (Chen *et al*., 2013) impacts of the science writing heuristic reported and negative and positive effects of collaboration during differentiation, depending on how data analysed (Eysink *et al*., 2017)
Summary of the evidence base	• Small number of studies • Study design: two randomised control trials • Two large-scale studies • Some inconsistency • Variability in interventions and outcome measures over studies as a whole. Few studies used standardised measures; three studies used researcher-developed instruments • All interventions were developed and evaluated by the researchers reporting the work

B4 Critical thinking and argumentation

Definitions

Critical thinking and scientific reasoning are related constructs that include various types of higher-order cognitive and metacognitive processes. According to Dowd *et al.* (2018), critical thinking is broader and includes a range of processes and dispositions which can be applied differently in different disciplines and in everyday life, with scientific reasoning defined as a type of critical thinking involved in science. Connected to critical thinking, argumentation is understood to be a type of discourse central to science which involves the ability to build a justified relationship between a claim and its supporting evidence or data (Osborne *et al*., 2016). Work on argumentation in science education frequently uses adaptations of Toulmin's Argument Pattern which illustrates how an argument works in terms of claims, data, warrants, backing and rebuttals (e.g. Erduran *et al.*, 2004). There are strong connections between this cluster and the cooperative and collaborative (B3) and inquiry clusters (B9), as several studies investigated argumentation and reasoning in the context of both writing (Hand *et al*., 2013; Reeves *et al.*, 2013; Chen *et al.*, 2013; Kara & Kingir, 2021) and inquiry (Chen & She, 2015; Chen *et al.*, 2019).

Description of the evidence base

Table 11 summarises the key features of each study in the critical thinking and argumentation cluster.

Six studies investigated the effect of critical thinking and argumentation approaches on pupil outcomes.

Approaches included reasoning or argumentation-driven approaches (Arias *et al*., 2017; Chen *et al.*, 2016; Miller *et al.*, 2014; Tsai *et al.*, 2012), and an approach to teaching based on the integration of analytical, creative and practical skills (Sternberg et *al.*, 2014). Approaches in this cluster have a range of theoretical underpinnings. Social constructivism is the basis of a number of approaches (Arias *et al.*, 2017; Tsai *et al.*, 2012), along with the theory of successful intelligence (Sternberg *et al*., 2014), the cognitive reconstruction of knowledge model (Miller *et al.*, 2017) and a model of situational interest (Chen *et al.,* 2016).

The studies in this cluster were conducted with children aged 8-11 in a range of science topic areas. The studies were conducted in England, Taiwan, Turkey, and the USA.

All six studies used a quasi-experimental design. Using the quality assessment tool, five studies were assessed as moderate-quality and one as high-quality.

Three studies reported impacts on attainment and conceptual understanding, one reported data on attitudes, and four reported on critical thinking and argumentation. Measures included bespoke tests (Kara & Kingir, 2021), assessment rubrics (Sternberg *et al.*, 2014), and questionnaires (Chen *et al*., 2016).

Table 11: Key features of studies on critical thinking and argumentation on primary science teaching

Authors	**Country**	**Design**	**Age of pupils**	**Sample** pupils *schools*	**Approach and assessment**	**Comparison**	**Outcomes measured**	**Summary of effects**
Arias *et al.* (2017)	USA	Quasi-experiment	8-11	1152 *not stated*	Kit-based inquiry units with educative features on making predictions with justifications (electricity, ecosystems) QA result: moderate	Kit-based inquiry units without educative features	Critical thinking: researcher assessment of pupils' predictions	*Critical thinking* ES(R)=0.27
Chen *et al.* (2016)	Taiwan	Quasi-experiment	9-10	72 *1*	Modified Argument Driven Inquiry (ADI) (physical science topics) QA result: moderate	Teaching as usual (textbooks, teacher presentations, demonstrations and practical work)	Attitudes: researcher-developed measure for science learning engagement based on Kind, Jones, & Barmby, (2007) Critical thinking: researcher-developed measure of argumentation abilities	*Attitudes* ES(R): η^2=0.06 for engagement in learning science and anxiety in learning science *Critical thinking* ES(R): η^2=0.13
Kara & Kingir (2021)	Turkey	Quasi-experiment	9-10	107 *1*	Model-based Science Writing Heuristic (various topics) QA result: moderate	Teaching as usual of the same units	Attainment: concept tests developed by researcher Critical thinking: rubric for scoring written argumentation developed by researcher	*Attainment* Unit 1 ES(C): *g*= 1.20 Unit 2 ES(C): *g*=1.01 *Critical thinking* ES(R): η^2=0.91
Miller *et al.* (2014)	USA	Quasi-experiment	9-10	130 *4*	Argumentative discussion (shape of the Earth) QA result: moderate	Reading for no stated purpose or to prepare for a regular classroom discussion	Attainment: interview scored according to rubric designed by Vosniadou and Brewer (1992)	*Attainment* data presented do not facilitate calculation of ES.
Sternberg *et al.* (2014)	USA	Quasi-experiment	9-10	7702 *113*	Units of instruction based upon the theory of successful intelligence (light, magnetism) QA result: high	Teaching as usual (weak control), memory instruction (strong control), critical-thinking instruction (strong control)	Attainment: researcher designed rubric specific for the unit	*Attainment* No meaningful effect
Tsai *et al.* (2012)	Taiwan	Quasi-experiment	10-11	189 *1*	Cognitive apprenticeship web-based argumentation (CAWA) system (on vision) QA result: moderate	Arguments on paper	Critical thinking: researcher-designed assessment of argumentation tasks	*Critical thinking* ES(R): η^2=0.30 (daily life) ES(R): η^2=0.23 (vision)

Findings

Table 11 presents an overview of the effects of approaches in the critical thinking and argumentation cluster. Most studies in this cluster reported positive impacts across the range of outcomes measured, with only Sternberg *et al.* (2014) having mixed outcomes.

Attainment (knowledge and conceptual understanding)

Three studies reported effects of approaches involving critical thinking and argumentation on attainment and conceptual understanding. These studies used quite distinct approaches: the science writing heuristic (Kara & Kingir, 2021), preparing pupils for argumentative discussion (Miller *et al*., 2014) and the comparison of teaching units based on successful intelligence, memory or critical thinking (Sternberg *et al*., 2014). Effects were mixed. The small-scale study by Kara and Kingir (2021) reported a large positive effect of the science writing heuristic on attainment tests, and Miller *et al.* (2014) found that reading with the purpose of argumentative discussion was associated with conceptual change for some pupils (those holding multiple ideas about the shape of the Earth). In contrast, Sternberg *et al*. (2014) found weak but statistically significant results in favour of the unit based on successful intelligence (for teaching light) and critical thinking (for teaching magnetism). These mixed findings are consistent with the approaches in the previous cluster looking at collaboration during argumentation.

Attitudes, interest and motivation

One study focused on attitudinal outcomes, with Chen *et al*. (2016) reporting on a small quasi-experiment in a single school. They reported a medium positive effect on engagement and on reducing anxiety for pupils using modified argument-driven inquiry. Given the limited size and the context, caution must be applied in drawing conclusions from this study alone.

Critical thinking and argumentation

Four studies reported positive effects on critical thinking and argumentation. These involved the modified argument driven inquiry (Chen *et al*., 2016), the model-based science writing heuristic (Kara & Kingir, 2021) and the use of resources to support argumentation. Resources included educative materials (Arias *et al.*, 2017) and a web-based argumentation system based on the Toulmin Argument Pattern comprising claims, data, warrants, backings, rebuttals and qualifiers (Tsai *et al.*, 2012).

The approach reported by Arias *et al.* (2017) involved prompts for teachers to encourage pupils to justify their predictions. Written predictions that included justifications were significantly greater in number than those in the control group, as were the number of clear, aligned, and accurate justifications. Similarly, Tsai *et al.* (2012) found gains in argumentation for pupils using an online argumentation system who received immediate teacher feedback on their arguments, scaffolded prompts for when they got stuck and were provided with a reflection area designed to refine thinking compared to those who used the same web-based argumentation system without the feedback, prompts and reflection area. Both groups demonstrated better argumentation performance than pupils in the control group who worked on pen and paper and had limited or no access to modelling, coaching, scaffolding and feedback from teachers or peers. The small-scale studies of modified argument-driven inquiry (Chen *et al.*, 2016) and the science writing heuristic (Kara & Kingir, 2021) also reported gains in argumentation overall in the experimental group compared to the control group. These studies together indicate that prompts and feedback have a role to play in promoting argumentation.

Implementation and resources

Barriers to implementation are likely to include the time required for teacher professional development and to produce educative features in curriculum materials.

One study (Arias *et al.*, 2017) involved commercially available kits, which have an associated cost.

The biggest resource required is time on the timetable for science lessons to carry out the interventions in a meaningful way. The nature of teaching approaches characterised by critical thinking and argumentation requires time to support pupils' independent and sustained thinking to create justified predictions and explanations, perhaps more so than approaches that focus on uncritical knowledge acquisition.

Other barriers include frequent and extended use of ICT and associated specialist software associated with the cognitive apprenticeship web-based argumentation system. One-off costs include hardware and software and ongoing costs include virus protection, connection to the internet, recharging costs for class/school wide individual devices, maintenance, upgrades and technical support.

Estimation of the required resources is difficult to include as costs were not reported in the studies, although five reported funding, suggesting some costs associated with these innovations.

Gaps in the evidence base

A specific focus on Key Stage 1 pupils would be of value given the absence of studies including children of this age in this cluster. Only one study reports on attitudinal-related outcomes. Other relevant outcomes such as working scientifically did not feature as outcomes of studies in this cluster.

Critical thinking and argumentation are central to approaches popular in primary education such as philosophy for children (P4C) (Trickey & Topping, 2007) and cognitive acceleration in science education (CASE) (e.g. Adey, 2005), but no studies were found testing P4C or CASE in primary science education.

Quality of the evidence base

In this cluster, one study was assessed to be high quality, with the remaining five assessed to be moderate-quality. There are no RCTs in this cluster, although there are two larger scale quasi-experimental studies. Studies were conducted in primary classrooms so there is generalisability to routine practice. Measured outcomes tended to be closely aligned with the approach.

There is a high degree of inconsistency with approach, theoretical underpinning, measures used and outcomes of interest.

Applicability of the evidence base

All interventions were applied at the classroom level in schools. Interventions using a resource (such as the science writing heuristic) or discussion activity which could be implemented flexibly in lessons are likely to be applicable in England, whereas those involving more extended schemes of work may not be consistent with the aims of school curricula in England.

One study (Sternberg *et al.*, 2014) used strong and weak control groups as comparators for the intervention, albeit not for science outcomes. This approach is not widely used in studies in the review, but is likely to be useful in understanding the mechanisms by which interventions work.

Critical thinking was identified as important in terms of approach and outcome by teachers during the review of practice, particularly in relation to the desirability of helping pupils to think like a scientist. There are no reasons to think that the findings identified by the studies in this cluster are not replicable to English classrooms. Given the findings reported outcomes for pupils aged 8-11 years of age, applicability is limited to Key Stage 2 pupils.

Overall evidence statement/Key findings

The studies in this cluster indicate that integration of teaching approaches which engage pupils in critical thinking and argumentation have a positive effect on attitudes and critical thinking, however, this is based on a small number of studies. In common with the collaborative approaches involving argumentation in B3, the evidence on impact on attainment is mixed, with positive or null effects on attainment, indicating at least no evidence of harm. Table 12 provides a summary of the evidence.

Table 12: Summary of the evidence on critical thinking and argumentation approaches

Number of studies	6
Assessed quality of studies	High: 1 Moderate: 5 Low: 0
Location of studies	Taiwan: 2 Turkey: 1 USA: 3
Outcomes of focus	Attainment (knowledge and conceptual understanding): 3 Attitudes, motivation and interest: 1 Critical thinking and argumentation: 4
Design	Quasi-experimental design: 6
Consistency of findings	Results were mixed when compared with business as usual. Some studies included elements where no significant difference was detected between experimental and control groups
Summary of the evidence base	● Small number of studies ● Study design: no randomised control trials ● Few large-scale studies ● Few studies assessed as high-quality ● Variability in interventions and outcome measures. Most studies developed their own instruments, rather than using standardised measures ● All interventions were developed and evaluated by the researchers reporting the work

B5 Explicit instruction and related approaches

Definitions

Explicit instruction and related approaches are based on an information processing model of learning. Explicit instruction has been defined as:

> A group of research-supported instructional behaviours used to design and deliver instruction that provides needed supports for successful learning through clarity of language and purpose, and reduction of cognitive load. It promotes active student engagement by requiring frequent and varied responses followed by appropriate affirmative and corrective feedback, and assists long-term retention through use of purposeful practice strategies. (Hughes *et al.*, 2017, p. 143)

Hughes at al. (2017) describe the following five key pillars of explicit instruction:

- Segmenting complex skills
- Drawing attention to important features through modelling or think-alouds
- Using physical, visual, and/or verbal prompts and gradually withdrawing these as students demonstrate accuracy and understanding
- Providing feedback opportunities
- Creating opportunities for purposeful practice.

Explicit instruction includes related approaches including Direct Instruction ('Big DI'), direct instruction ('little di'), explicit teaching and explicit direct instruction. Direct Instruction (Engelmann *et al.*, 1988) involves the use of scripted lessons, including curriculum content (what to teach) and pedagogy (how to teach), whereas explicit instruction is concerned only with pedagogy. Similarly, direct instruction refers to components of instruction used by effective teachers, defined as those whose students scored highest in assessments of, usually, English and Mathematics (Rosenshine, 2009). These components included daily practice, presenting new material in small steps and checking for understanding, guiding student practice (e.g. connecting with prior knowledge and summarising), providing feedback, independent practice, and weekly or monthly review.

Studies were assigned to the explicit instruction cluster where explicit (or direct) instruction or its key characteristics featured as approach.

Description of the evidence base

The *explicit instruction and related approaches* cluster comprises 14 studies. The summary of effects is presented in Table 13.

A variety of approaches are evaluated within this cluster. Explicit instruction was used within an inquiry curriculum context (Doabler *et al.*, 2021; Upadhyay & DeFrano, 2008; van der Graaf *et al.*, 2019; and see also Schalk *et al.*, 2019 in the inquiry cluster) - although inquiry was not always defined. It was also used in the context of vocabulary or reading comprehension (Williams *et al.*, 2009). Studies using pillars of explicit instruction (Hughes *et al.*, 2017) involved modelling, for example using instructional scaffolds such as adapted worksheets (Baumfalk *et al.*, 2019; Zangori *et al.*, 2015), the use of images (Cohen & Johnson, 2012; Yeo *et al.*, 2020), questions or prompts (Rotgans & Schmidt, 2017), multimedia and analogy (Zheng *et al.*, 2008) and toys and specimens (Randler, 2009). Metacognitive instruction (Michalsky *et al.*, 2009) and concept mastery (Kim *et al.*, 2012) were also included in this cluster.

Table 13: Key features of studies on explicit instruction and related approaches

Authors	Country	Design	Age of pupils	Sample pupils *schools*	Approach and assessment	Comparison	Outcomes measured	Summary of effects
Baumfalk *et al.* (2019)	USA	Quasi-experiment	8-9	201 *9*	Model-enhanced worksheets (water cycle) QA result: moderate	Full Option Science System unit without modelling enhanced worksheets	Working scientifically: modelling task scored using rubric designed by researchers	*Working scientifically* ES(R): *d*=1.12
Berry, Potter & Hollas (2013)	USA	Quasi-experiment	8-9	58 *1*	Concept maps pre- and post-reading activity, sharing with class (soil) QA result: low	Teacher questioning pre-reading, with pupils writing answers then sharing with classmates	Attainment: tests developed by the researchers	*Attainment* ES(C): *g*=1.08 (multiple choice test); ES(C): *g*=0.94 (writing task) ES(C): *g*=1.22 (relational knowledge)
Cohen & Johnson (2012)	USA	Quasi-experiment	10-11	89 *2*	Use of images to support science vocabulary learning QA result: moderate	Simple verbal presentation of the scientific term and concepts.	Scientific language: vocabulary comprehension tests designed by researchers.	*Scientific language* data presented do not facilitate calculation of ES
Doabler *et al. (*2021)	USA	Cluster RCT	7-8	291 *3*	Explicit instruction (guided inquiry) unit (Earth science) QA result: high	Teaching as usual (blend of district-developed materials and commercially available science programmes).	Attainment: Test of Early Geoscience Learning [TEGL] based on (Doabler, Longhi, Maddox, *et al.*, 2019) and Content Knowledge and Scientific Practices [CKSP] (Assessment Technology Incorporated [ATI], 2019). Scientific language: Science Vocabulary Knowledge [SEVA] Working scientifically: Virtual Interactive Scientific Practices Assessment [VISPA] based on Doabler, Longhi, Uy, *et al.*, 2019), and CKSP above.	*Attainment* TEGL ES(R): *g*=0.60 *Scientific language* SEVA, ES(C): *g*=0.94 *Working scientifically* VISPA ES(C): *g*=0.48 CKSP ES(C): *g*=0.02

Table 13 (continued)

Authors	Country	Design	Age of pupils	Sample pupils *schools*	Approach and assessment	Comparison	Outcomes measured	Summary of effects
Kim *et al.* (2012)	USA	Quasi-experiment	5-9	2182 *6*	Concept mastery and investigation unit QA result: low	School-based curricula	Attainment: MAT-8 science subtest) Critical thinking: standardised measure of critical thinking	*Attainment* ES(R): η^2=0.013 *Critical thinking* ES(R): η^2=0.03
Michalsky, Mevarech & Haibi (2009)	Israel	Quasi-experiment	9-10	108 *4*	Metacognitive instruction before, during or after a collaborative group task (animals and plants) QA result: moderate	No metacognitive instruction.	Attainment: test of scientific knowledge designed by Ministry of Education, test based on PISA scientific literacy test Critical thinking: adapted version of the Metacognition Awareness Questionnaire (Schraw & Dennison, 1994) Working scientifically: researcher designed scientific literacy test	*Attainment* ES(R): η^2=0.13 *Critical thinking* ES(R): η^2=0.34 *Working scientifically* ES(R): η^2=0.24
Randler (2009)	Germany	Quasi-experiment	6-10	138 *not stated*	Specimens (bird identification) and worksheets QA result: moderate	Soft toys and worksheets	Attainment: test on bird species identification designed by researcher	*Attainment* No difference between toy and specimen conditions
Rotgans & Schmidt (2017)	Singapore	Quasi-experiment	9-10	129 *1*	Situational interest inducing problems and goal setting (light) QA result: moderate	Provided with similar information by the teacher	Attitudes: individual interest and situational interest tests developed by researchers	*Attitudes* Data presented do not facilitate calculation of ES
Upadhyay & DeFranco (2008)	USA	Quasi-experiment	8-9	108 *2*	Connected science instruction (environmental science) QA result: low	Direct instruction (not defined)	Scientific language: researcher-developed environmental science survey	*Scientific language* ES(R): η^2=0.038 in favour of control group

Table 13 (continued)

Authors	Country	Design	Age of pupils	Sample pupils *schools*	Approach and assessment	Comparison	Outcomes measured	Summary of effects
van der Graaf *et al.* (2019)	Netherlands	RCT	9-10	301 *10*	Direct instruction combined with verbal support teacher training. Implemented in the context of scientific reasoning during a series of inquiry-based lessons (control of variables) QA result: moderate	Direct instruction only, verbal support only, lesson-series (inquiry) only (baseline condition).	Attainment: researcher-designed domain specific knowledge test Critical thinking: scientific reasoning inventory adapted by researchers Scientific language: domain-specific vocabulary test developed by researchers	*Attainment* ES(R): *d*=0.22 or better (different levels of transfer reported separately) *Critical thinking* ES(R): *d*=0.30 or better (different components reported separately) *Scientific language* ES(R): *d*=0.64
Williams *et al.* (2009)	USA	Quasi-experiment	7-8	215 *4*	Explicit instruction of scientific vocabulary (animals) QA result: moderate	Two comparisons: a content lesson programme group and a no instruction group	Scientific language: Woodcock reasoning mastery test pre-test and researcher developed tests	*Scientific language* ES(R): *d*=1.36-4.40 but data as presented do not facilitate detailed deductions
Yeo *et al.* (2020)	Singapore	Quasi-experiment	9-10	129 *2*	Image to writing inquiry the use of images to represent ideas, and translation of images into text (temperature and heat) QA result: low	Direct instruction and inquiry (using predict-observe-explain) without multimodal representations	Attainment: researcher designed coding framework for conceptual understanding Scientific language: researcher designed coding framework for representational competence	*Attainment* ES(R): *d*=0.42 *Scientific language* (representational competency) ES(C): *g*=0.46
Zangori, Forbes & Schwarz (2015)	USA	Quasi-experiment	8-9	116 *not stated*	Model enhanced worksheets (water cycle) QA result: low	Unscaffolded worksheets, non-equivalent group	Working scientifically: modelling task scored using rubric designed by researchers	*Working scientifically* data presented do not facilitate calculation of ES
Zheng *et al.* (2008)	USA	Quasi-experiment	9-10	89 *1*	Model-based reasoning (electricity) QA result: low	Three comparisons: multimedia with and without analogy, analogy without multimedia and instruction involving neither multimedia nor analogy	Attainment: Achievement test designed by researchers Critical thinking: Group Embedded Figure Test (Witkin *et al*., 1971; 2002)	*Attainment* ES(R): η^2=0.16 (for full dose intervention on recall) *Critical thinking* ES(R): η^2=0.16 (for full dose intervention on transfer)

The studies were conducted in a range of topic contexts across biology, chemistry, Earth science and physics as well as in teaching scientific practices. The age of children involved ranged from 5-11, although the majority of studies involved children of Key Stage 2 age (8-11), with only two involving children younger than 8 years old. Studies were conducted in Germany, Israel, the Netherlands, Singapore and the USA. No studies from the UK were included in this cluster.

A quasi-experimental design was used in 12 studies, with two using a randomised control trial design. Through a quality assessment instrument, one study was rated as high-quality, seven studies were rated as moderate-quality and six as low-quality.

Intended outcomes included achievement in science (Doabler *et al.*, 2021), interest (Rotgans & Schmidt, 2017), application of experimental design (Schalk *et al.*, 2019) knowledge, scientific literacy and metacognitive awareness (Michalsky *et al.*, 2010), and text comprehension (Williams *et al.*, 2009). A wide range of measures were used in this cluster. These included tests of science (Doabler *et al.*, 2021), critical thinking or reasoning (van der Graaf *et al.*, 2019), and vocabulary (Williams *et al.*, 2009). Other measures included an interest questionnaire (Rotgans & Schmidt, 2017). Instruments in this cluster tended to be bespoke instruments designed by the research teams, although standardised vocabulary and reasoning mastery tests (Cohen & Johnson, 2012; Williams *et al.*, 2009), and cognitive tests (Zheng *et al.*, 2008) were used in some studies, as were adapted versions of existing questionnaires (Michalsky *et al.*, 2010).

Findings

There was considerable diversity in the nature and duration of interventions, theoretical underpinnings, the science topic interventions were used in, the intended outcomes for children, and the measures used to assess the success of the intervention. Explicit instruction was often incorporated into inquiry curriculum contexts. The most common outcomes measured across the cluster were conceptual understanding (albeit in different topics), working scientifically and scientific language.

Most studies reported positive or positive or null outcomes resulting from the use of explicit instruction and related approaches.

Attainment (knowledge and conceptual understanding)

Eight studies reported impacts of explicit instruction and related approaches on attainment.

Three focused on explicit or direct instruction during inquiry. One cluster randomised control trial (Doabler *et al.*, 2021) investigated the impact of a 10x30 minute lesson 'Science Explorers' unit incorporating steps to activate prior knowledge and offer contact, introduce new vocabulary, read-aloud, investigate and share learning with families. Fundamental concepts and skills were explicitly taught. The study found a medium positive effect on geoscience learning and small and non-significant positive effects on content knowledge. This positive outcome is supported by the positive effect reported in the smaller scale study of direct instruction on the control of variable strategy during inquiry (van der Graaf *et al.*, 2019), and by the study by Kim *et al.* (2012) who found small effects of teaching using units which integrate concept mastery using higher-level questions, reflection and discussion, into an inquiry unit.

One study focused on a metacognitive approach. Reading scientific texts with metacognitive instruction was found to be more effective than instruction without metacognitive intervention across a range of measures including scientific knowledge and scientific literacy (Michalsky *et al.*, 2009). Order matters: those who had metacognitive instruction *after* reading the scientific text significantly outperformed all groups.

Four studies focused on the use of images, diagrams and models. Yeo *et al*. (2020) investigated an approach to learning which involved representing ideas as images and then converting images to text. They reported medium positive effects on conceptual understanding. Zheng *et al.* (2008) investigated the use of analogical reasoning in a brief intervention and found small gains in recall and transfer of knowledge between contexts (water and electrical circuits), although different constructs were measured pre- and post-intervention. Berry, Potter and Hollas (2013) investigated the impact of concept mapping on attainment in a small-scale quasi-experiment and found positive effects on relational knowledge. Finally, Randler (2009) found no differences between the use of toys and taxonomic specimens in the teaching of bird species.

In terms of outcomes for specific groups of children, Doabler *et al.* (2021) found that the science explorers programme is promising for improving outcomes for children with, or at risk of developing, difficulties with learning, measured using a knowledge pre-test.

Attitudes, interest and motivation

One study collected data on attitudes and interest. Rotgans and Schmidt (2017) investigated the impact of problem posing on inducing interest in science. They reported gains in interest for pupils who experienced problem posing and goal setting, and loss of interest for pupils in the group which was provided information by the teacher. The effect size could not be determined from the data presented.

Critical thinking and argumentation

Four studies reported positive effects on critical thinking. The approaches included direct instruction during inquiry (van der Graaf *et al.*, 2019), concept mastery during investigation (Kim *et al.*, 2012), model-based reasoning (Zheng *et al.*, 2008) and metacognitive instruction (Michalsky *et al.*, 2009). The metacognitive, mastery and direct instruction during inquiry interventions had larger effects on critical thinking than they did on attainment.

Scientific language and communication

Six studies were concerned with language and communication outcomes.

Of four studies involving explicit instruction, the cluster randomised control trial of a guided inquiry 'Science Explorers' unit involving explicit teaching of scientific vocabulary (Doabler *et al.*, 2021) and the quasi-experimental study of an intervention involving explicit instruction in text structure (Williams *et al.,* 2009) both reported a large positive effect on science vocabulary test outcomes.

The randomised control trial comparing the teaching of inquiry using direct instruction, verbal support from the teacher, and a combination of direct instruction and verbal support with a baseline condition (van der Graaf *et al.,* 2019) found that only the combined direct instruction and verbal support approach had an effect on vocabulary outcomes. However in this case, the direct instruction referred to the control of variables strategy, whereas in the studies by Doabler *et al.* (2021) and Williams *et al.*, 2009 (above) the vocabulary that was tested was explicitly taught. Finally, Upadhyay & DeFranco (2008) compared connected science instruction with direct instruction as the control group. Whilst direct instruction was not defined, they found that pupils in this group gained more vocabulary (but lost it at a greater rate) than pupils in the connected science instruction group.

Two studies focused on connecting images or diagrams to writing. Cohen and Johnson (2012) reported significant gains in retention and comprehension of scientific vocabulary when pupils were asked to create and draw an image relating to vocabulary taught compared with pupils who were presented with a word or an image. However, the intervention appears to have taken place out of context, i.e. words not necessarily related to the science being taught as part of a science unit. Finally, a quasi-experiment investigating image-to-writing (Yeo *et al.*, 2020) found a medium effect on scientific language competence compared with the control group which did not focus on multimodal representations.

Working scientifically and problem solving

Three studies reported on outcomes relating to working scientifically and problem solving. One of these was the randomised control trial of a Science Explorers unit which incorporated explicit instruction into an enquiry unit (Doabler *et al.*, 2021) and two measured the impact of a modelling enhanced curriculum (Baumfalk *et al.*, 2019; Zangori *et al.*, 2015). The two studies on the modelling enhanced curriculum involved some of the same authors. The study which tested the effects of embedding explicit instruction into a guided inquiry based curriculum (Doabler *et al.*, 2021) found a small positive effect on virtual interactive science practices and small and non-significant positive effects on scientific practices.

A study on a modelling enhanced curriculum (Baumfalk *et al.*, 2019) found that pupils in the intervention group had more opportunities to engage with models and were better able to emphasise non visible components of the water cycle (e.g. groundwater), identify greater numbers of sequences, and use components such as the sun to explain processes such as evaporation.

Implementation and resources

No major barriers to implementation were reported. However, several studies involved the design of instructional units or bespoke resources and/or professional development for teachers (e.g. Baumfaulk *et al.*, 2019; Doabler *et al.*, 2021), which has implications for both cost and practicality of implementation.

The curriculum developed by Doabler *et al.* (2021) was delivered daily over a short time period (two weeks). It would be unusual for children in primary schools in England to have a science lesson every day, but it is possible, and the curriculum could be delivered over a longer period, although it is not known whether intensity of delivery might be an important factor in the positive outcomes.

Gaps in the evidence base

Few studies in the explicit instruction and related approaches cluster discussed the impact of interventions on attitudinal outcomes. Only Rotgans & Schmidt (2017) focused on interest, and none of the studies focused on confidence, self-efficacy, or participation in or access to science.

Quality of the evidence base

Using the quality assessment tool, one study was judged to be of high-quality, seven moderate-quality and six low-quality. There were few large-scale studies involving pupils from a large number of schools in this cluster. All studies were conducted in classrooms, with good generalisability to routine practice. Whilst there are a larger number of studies than in some clusters, there is a high degree of heterogeneity in approach and outcome measure, with small numbers of studies using robust designs in which researchers are independent of the intervention and use standardised assessment methods.

Proximal measures, i.e. those close to the intervention, featured in this group of studies. These can be problematic when assessing the effectiveness of an intervention if the control group is not taught the same topic or skill being assessed.

Applicability of the evidence base

None of the approaches in this cluster were conducted in England, although the interventions are broadly applicable to the English context as they involved interventions that were delivered by teachers, e.g. curriculum units or teaching activities.

Costs were not reported, although funders of studies in this cluster included the National Science Foundation in the USA (Doabler *et al.*, 2021; Baulmfalk *et al.*, 2019), the Netherlands Initiative for Educational Research (van der Graaf *et al.*, 2019) and the National Institute of Education in Singapore (Yeo e*t al.*, 2020). Costs of implementation may be low for delivery if curriculum materials are provided at no cost to schools. However there are costs associated with research-informed curriculum development and professional development of teachers which are likely to be more considerable. There is considerable expertise in close-to-practice research in primary science, and evidence-informed curriculum development in science education in England.

Overall evidence statement/Key findings

Evidence suggests that explicit instruction and related approaches have some positive effects across a range of outcomes including attainment, critical thinking, scientific language and working scientifically, most notably in the assessment of targets of the explicit instruction, i.e. in teaching scientific vocabulary and the control of variables. A summary of the evidence is presented in Table 14.

Table 14: Summary of the evidence on explicit instruction and related approaches

Number of studies	14
Assessed quality of studies	High: 1 Moderate: 7 Low: 6
Location of studies	Germany: 1 Israel: 1 Netherlands: 1 Singapore: 2 USA: 9
Outcomes of focus	Attainment (knowledge and conceptual understanding): 8 Attitudes, motivation and interest: 1 Critical thinking and argumentation: 4 Scientific language: 6 Working scientifically and problem solving: 4
Design	Quasi-experimental design: 12 Randomised control trial: 2
Consistency of findings	There was a diverse set of outcomes measured using a wide variety of instruments. Results were consistently positive across a range of measures for explicit instruction and modelling strategies when compared with business-as-usual
Summary of the evidence base	• Study design: two randomised control trials • Few studies with large sample sizes • Variability in interventions and outcome measures • Outcome measures were often designed by researchers and very closely aligned to intervention approach and target area • Few studies used standardised instruments

B6 ICT supported and online teaching and learning

Definitions

Information Communication Technology (ICT) can be defined as "a diverse set of technological tools and resources used to transmit, store, create, share or exchange information. These technological tools and resources include computers, the Internet (websites, blogs and emails), live broadcasting technologies (radio, television and webcasting), recorded broadcasting technologies (podcasting, audio and video players, and storage devices) and telephony (e.g. fixed or mobile, satellite, visio/video-conferencing)" (UNESCO, 2009). Although there is debate about how computing can - or should - be integrated into the wider curriculum (e.g. McGarr & Johnston, 2020), and although none of the studies in this section were conducted in the UK, of most relevance to note here is the explicit connection of the computing curriculum with primary science. The ways in which ICT can be used to enhance scientific knowledge, understanding and skills is of primary interest in this section.

Description of the evidence base

Table 15 summarises the key features of the ten studies reporting on ICT supported and online teaching and learning approaches.

Four studies focused on models or immersive experiences otherwise difficult to re-create in the classroom: animated movies (Barak & Dori, 2011); serious educational gameplay (Hodges *et al.*, 2020); 3D Virtual reality models (Sun *et al.*, 2010); virtual materials and simulations of physical resources (Wang &Tseng, 2018). Four studies focused on the online presentation of materials, working spaces and recording of work: questioning-supported thinking and learning system (Hu *et al.*, 2019); a concept map-guided problem-posing tool for flipped learning (Hwang *et al.*, 2020); a concept map based electronic portfolio (Kim & Olaciregui, 2008); multimedia online tutor (Ward *et al.*, 2013). Two studies focused on individual electronic devices: a mobile learning environment comprising a smartphone with apps connected to the internet (Looi *et al.*, 2011); mobile devices as a means of data collection (Zacharia *et al.*, 2016).

The studies were conducted in various topics across biology, chemistry, physics and Earth science, with children aged 8-11. Four studies were conducted in Taiwan, three in the USA and one each in Cyprus, Israel and Singapore.

One study adopted a randomised control trial design and the remaining nine were quasi-experimental studies. Nine were assessed to be moderate- and one was assessed to be of low-quality using the quality assessment tool.

All of the studies embedded the use of digital technology in a science lesson as a pedagogical tool and reported on science specific outcomes. Nine reported on attainment, with three also reporting on attitudinal outcomes and two reporting on critical thinking. Measures included researcher- and teacher-designed tests and questionnaires, and previously validated critical thinking tests and self-efficacy scales.

Findings

Where effects were reported or calculated, these tended to be positive or on a par with the control group. We caution against ranking or comparing approaches by effect sizes given that few studies use similar or standardised measures, or target the same construct.

Table 15: Key features of studies on ICT supported and online in primary science teaching

Authors	**Country**	**Design**	**Age of pupils**	**Sample** pupils *schools*	**Approach and assessment**	**Comparison**	**Outcomes measured**	**Summary of effects**
Barak & Dori (2011)	Israel	Quasi-experiment	9-11	1335 *7*	Integration of animated movies (various topics) QA result: moderate	Textbooks and still images	Critical thinking: science thinking skills questionnaire	*Critical thinking* ES(R): η^2=0.22
Hodges *et al.* (2020)	USA	Quasi-experiment	8-11	232 *1*	Serious game and guided inquiry (animal body systems) QA result: moderate	Card sort and hands-on activities, same topic	Attainment: Test aligned with content taught to both groups designed by researchers	*Attainment* ES(C): g=0.36
Hu *et al.* (2019)	Taiwan	Quasi-experiment	10-11	100 *1*	Online questioning system on rainfall:detailed or simple question stem QA result: moderate	Online questioning system on rainfall: no question stem	Attainment: achievement quiz designed by teachers, teacher assessment of reports Critical thinking: Critical Thinking Test-Level I developed by Yeh (2003), classification of question types	*Attainment* No significant effect (achievement quiz); ES(R): η^2=0.11 (teacher scoring of reports) *Critical thinking* ES(R): η^2=0.07
Hwang *et al.* (2020)	Taiwan	Quasi-experiment	10-11	75 *1*	Flipped learning with problem posing (plants) QA result: moderate	Worksheet. All groups received the same materials and completed an activity at home	Attainment: test designed by teachers Attitudes: self-efficacy scale (Pintrich *et al.*, 1991)	*Attainment* ES(R): η^2=0.11 (for full dose intervention) *Attitudes* no significant effect
Kim & Olaciregui (2008)	USA	RCT	10-11	50 *1*	Concept mapping information display in electronic portfolio (atmosphere) QA result: moderate	Tree mode information display in electronic portfolio	Attainment: multiple choice and comprehension test	*Attainment* ES(C): *g*=1.39
Looi *et al.* (2011)	Singapore	Quasi-experiment	8-9	351 *1*	Mobile technology smartphone apps) in inquiry context (body systems) QA result: low	'Teaching as usual'	Attainment: general science examination exam Attitudes: attitudes towards mobile devices questionnaire	*Attainment* ES(R): η^2=0.41 *Attitudes*: data presented do not facilitate calculation of ES

Table 15 (continued)

Authors	Country	Design	Age of pupils	Sample pupils *schools*	Approach and assessment	Comparison	Outcomes measured	Summary of effects
Sun *et al.* (2010)	Taiwan	Quasi-experiment	10-11	128 *1*	Virtual reality modelling (Sun-Moon-Earth system) QA result: moderate	2D photographs	Attainment: conceptual test developed by the researchers Attitudes: attitude questionnaire based on Sun, Lin and Yu (2008)	*Attainment* ES(C): g=0.34 *Attitudes* data presented do not facilitate calculation of ES
Wang & Tseng (2018)	Taiwan	Quasi-experiment	8-9	208 *1*	Virtual interactive laboratory before physical manipulation (changes of state) QA result: moderate	Physical manipulation activities on same topic	Attainment: science achievement test developed by researchers and conceptual test based on Chang (2002)	*Attainment* ES(R): η^2=0.062 (achievement test)
Ward *et al.* (2013)	USA	Quasi-experiment	8-11	1167 *22*	Virtual tutor (MyST) or human tutor assigned (various topics) QA result: moderate	No supplementary tutoring, same topics in class	Attainment: Assessing Science Knowledge (ASK) assessments from Full Option Science System programme	*Attainment* ES(R): d=0.53 (MyST) ES(R): d=0.68 (human tutoring)
Zacharia *et al.* (2016)	Cyprus	Quasi-experiment	9	48 *1*	Mobile devices for data collection QA result: moderate	Equipment and books for data collection	Attainment: conceptual knowledge test	*Attainment* ES(R): η^2=0.30

Attainment (knowledge and conceptual understanding)

Nine studies reported impacts of ICT and online approaches on attainment, with effect sizes reported or calculated for each.

Three studies that used models or immersive experiences otherwise difficult to re-create in the classroom on the whole suggest a small to medium positive effect on pupil attainment. These included serious educational gameplay (SEG) in the context of a Virtual Vet (Hodges *et al.*, 2020), an ICT application (Sun *et al.*, 2010) showing the relative positions and movements of the sun and moon with use of a 3-D VR model to assist night-time observations, and a virtual laboratory to explore the processes of evaporation and condensation (Wang & Tseng, 2018). These were compared with textbooks and still pictures, use of physical material (e.g. to build a model of a cat's leg, dissection of an owl pellet and a card sorting activity) and the support of 2D photographs.

Four studies focused on online presentation of materials, working spaces and recording of work. Three (Hwang *et al.*, 2020; Kim & Olaciregui, 2008 and Ward *et al.*, 2013) reported positive effects on attainment and Hu *et al.* (2019) found no meaningful effect of a questioning-supported thinking and learning system on achievement. Positive effects were reported as a result of using a concept map guided problem-posing tool (Hwang *et al.*, 2020), a concept map based electronic portfolio (Kim & Olaciregui, 2008) and an online tutor (Ward *et al.*, 2013). Hwang *et al.* (2020) and Kim and Olaciregui (2008) identified how accessing or recording information in a concept map layout led to better pupil outcomes compared with alternative online layouts or paper-based alternatives. Kim and Olaciregui (2008) found a positive effect of using a concept-map interface compared with using the conventional tree mode content organisation, with faster searching and higher retention accuracy. In the study by Hwang *et al.* (2021), the ICT functionality allowed the embedding of a multilevel problem-posing strategy mechanism into the programme enabling the teacher to monitor pupil progress through a management and feedback mechanism (although it was not clear whether the pupils read and integrated any teacher feedback into their responses). It is also important to note that the pupils in the intervention group had higher measures of critical thinking tendency at pre-test.

Two studies focused on the use of mobile equipment outside the classroom for data collection and recording, with small to medium effects on attainment. Looi *et al.* (2011) investigated the impact of the use of bespoke and flexible learning tools on mobile phones as part of an inquiry curriculum. They found a positive effect on general science examination scores compared with the group which experienced teaching using textbooks and teacher-led 'practice and drill' sessions or by using physical tools such as magnifying glasses and sketchbooks. Zacharia *et al.* (2016) investigated the use of mobile devices to collect data about flowers on a field trip. They reported a positive effect on conceptual understanding and scientific accuracy when compared with sketching and note-taking.

Attitudes, interest and motivation

Three studies investigated the impacts of ICT supported and online teaching on attitudes and related outcomes, although the results are inconclusive.

Looi *et al.* (2011) noted a shift in pupil behaviour after the introduction of the smartphones, with 62% of the experimental group reporting that the phones had helped them to understand science concepts better and to understand how their learning was connected to daily life.

One goal of the study by Hodges *et al.* (2020) was to find out if a serious educational game could alter pupil mindsets (as defined by Dweck, 2006). They observed an interaction effect that pupils with a growth mindset (a belief that gains can be made through effort and perseverance) collectively experienced more growth from pre-to post-intervention compared to pupils characterised by a more fixed mindset (a belief that intelligence and ability cannot change).

Hwang *et al.* (2020) found no significant differences in self-efficacy between the different groups involved in the study exploring the use of a concept map problem posing strategy during flipped learning.

Critical thinking and argumentation

Two studies reported on critical thinking outcomes. These studies investigated the impact of the use of animated movies (Barak & Dori, 2011) and the use of an online questioning system (Hu *et al.*, 2019).

In terms of prompts and resources, Barak and Dori's (2011) findings indicated a small effect on thinking skills as a result of using animated movies compared with the use of textbooks and still images. Hu *et al.* (2019), in their assessment of critical thinking in a collaborative online questioning system, found that a group given a detailed question stem or no question stem at all performed significantly better than groups given a simple question stem.

The claim that ICT and online learning is an effective approach in improving pupil outcomes is supported by the following factors identified in the findings from the studies in this cluster including immediate feedback, opportunities for collaboration, multiple representations of concepts and ideas, learning tailored to the individual, pupil autonomy in the choice of ICT tools and ways of recording, which can be revisited, providing virtual experiences (e.g. models and simulations) of contexts or ideas that cannot readily be recreated in the classroom.

Implementation and resources

Although few studies referenced implementation and resource challenges, there are some barriers to the integration of ICT into science teaching and learning relating to cost, access to technology and confidence with technology.

Some of the resources (e.g. *Brainpop*) have a cost associated with their use, and researchers or developers in some of the studies provided professional development and support in using the ICT resources (e.g. Barak & Dori, 2011; Hodges *et al.*, 2020). Estimation of the required resources is difficult to include as costs were not reported in the studies and these are likely to differ across country and time. As previously noted in B4, costs associated with ICT and online approaches include hardware and software costs, service costs, charging costs, and technical support. Rapid obsolescence of hardware and software means that the investment may not be sustained for a long period of time.

It is also important to consider costs to families where the expectation is that pupils use their own devices or where they are expected to learn at home using a prescribed system as in the case of the study of flipped learning using a problem posing mechanism (Hwang *et al.*, 2020). The lockdowns caused by the COVID-19 pandemic highlighted a digital divide that had implications for educational progress when schools moved to online teaching and learning. An Ofcom survey carried out between January and March 2020 found that 9% of households containing children did not have home access to a laptop, a desktop PC or tablet. Another study carried out by the UCL Institute of Education (Green, 2020) identified that one in five children eligible for free school meals had no access to a computer at home. Although children of key workers and those deemed vulnerable had the opportunity to attend school during the lockdown, the ongoing nature of the inequalities associated with access to technology and connectivity to the internet have implications for children's learning that is supported by homework tasks.

Other barriers relate to teacher confidence in the use of technology and the cost and time for related CPD and implementation of online safeguarding measures. Barak and Dori (2011) highlight that, prior to intervention, 20% of teachers reported that they did not use computers at all. Following the intervention, they reflected that they had become more technologically and pedagogically savvy.

Gaps in the evidence base

There were no studies that included Key Stage 1 aged pupils, and no studies conducted in England. Few studies reported on attitudinal outcomes or critical thinking, and none reported impacts on pupils' ability to work scientifically. It would be useful to assess the impact on learning as a result of the use of recorded lessons (for example, as provided by Oak Academy) and online tools which became prominent during the pandemic. Hodges *et al.* (2020) note that technology infused lessons are replacing other learning experiences. As for any pedagogical tool, it is important that the teacher understands its purpose and how their use - whether on their own or combined with other approaches - can lead to better pupil outcomes.

A future focus on how ICT can support pupils to develop scientific conceptual knowledge (and challenge common misconceptions) would provide useful evidence on which to build targeted support for pupils.

Four of the ten studies in this cluster are from Taiwan which prioritises funding for ICT in schools. In 2013, The Taiwanese Ministry of Education launched a major e-learning initiative where wireless capability was expanded in schools so that pupils in every school would be able to use broadband networks on campus (NCEE, 2022). Despite this, a digital divide was identified by The Child Welfare League Foundation (CWLF, 2022) which warned of increased educational inequality in Taiwan due to COVID-19 with almost 70 percent of rural school goers reporting that they did not own a device appropriate for online classes during the lockdown in May 2021. It is important to consider the implications of any technological innovations on children who do not have access to technology at home.

Quality of the evidence base

Of the ten studies included in this cluster, nine were assessed to be moderate- and one was assessed to be low-quality. Only one study (a relatively small one) used a randomised control trial design.

All studies were applied at the classroom level and conducted in typical learning environments including classrooms, ICT rooms or outside, and learning was led by a teacher. Only two studies were large-scale.

This cluster included a wide range of approaches, targeting different outcomes and using different measures.

Applicability of the evidence base

Although the education systems are different in the countries where the studies were conducted, based on the age groups, educational settings, classroom organisation, teacher confidence and science curriculum topics reported in these studies, there is no evidence to suggest that these findings would not hold in England if it could be ensured that devices and applications were made available to children.

Overall evidence statement/Key findings

Findings are consistent across studies, with studies using ICT supported or online approaches typically reporting positive effects on attainment. However, the approaches and outcomes measured vary considerably. Evidence is inconsistent for the effect on attitudinal outcomes and critical thinking, with fewer studies reporting on these outcomes. The evidence is summarised in Table 16.

Table 16: Summary of the evidence on ICT-supported and online teaching and learning

Number of studies	10
Assessed quality of studies	High: 0 Moderate: 9 Low: 1
Location of studies	Cyprus: 1 Israel: 1 Singapore: 1 Taiwan: 4 USA: 3
Outcomes of focus	Attainment (knowledge and conceptual understanding): 9 Attitudes, motivation and interest: 3 Critical thinking and argumentation: 2
Design	Quasi-experimental design: 9 Randomised control trial: 1
Consistency of findings	Most ICT supported and online approaches reported a positive or null effect when compared with business-as-usual.
Summary of the evidence base	• Study design: one randomised control trial • Lack of high-quality studies and studies with large sample sizes • Several studies conducted in single schools • Most interventions were developed and evaluated by the researchers reporting the work • Variability in interventions and outcome measures. Most studies developed their own instruments, rather than using standardised measures

B7 Language, literacy and text-based approaches

Definitions

Language, literacy and text-based approaches refers to teaching approaches which involve aspects of literacy e.g. reading, writing and talking about science. It also includes approaches which focus on science teaching for pupils who speak English as an Additional Language (EAL). In the UK, pupils who speak EAL account for around 20% of the school population (School Census Statistics Team, 2022).

Description of the evidence base

The *language and literacy* cluster comprises 18 studies. Alexander (2018) and Jay *et al.* (2017) include data in common (with Alexander drawing on a wider range of data sources), as do Hanley *et al.* (2015) and Hanley *et al.* (2020), and Cervetti *et al.* (2012) and Goldschmidt and Jung (2011).

Table 17 summarises the key features of each study. A variety of language, literacy and text-based approaches are evaluated in the studies. These include science and literacy integrated curriculum units, some with a focus on English language learners (Llosa *et al.*, 2016; Maerten-Rivera *et al.*, 2016), approaches to classroom talk involving substantial CPD such as dialogic teaching (Alexander, 2018; Jay *et al.*, 2017) and Thinking, Doing, Talking Science (Hanley *et al.*, 2015;2020; Kitmitto *et al.*, 2018), as well as reading, writing and textbook-based interventions.

The age of the pupils involved spans the whole range as represented in UK primary schools (4-11), with four-year-olds only included alongside their older peers, and the majority focusing on older primary-aged pupils. The majority of studies in this cluster were conducted in the USA (9), with others taking place in the UK (5), Taiwan (2), the Netherlands (1) and Italy (1). The approaches were used in a variety of science topic contexts across biology, chemistry, Earth science and physics.

Twelve studies adopted a randomised control trial design (with varying levels of robustness, size, stratification) and six used a quasi-experimental design. Twelve were rated high-, four moderate- and the other two were low-quality as judged by the quality assessment tool. Control conditions varied. Some control teachers were asked to do their teaching as usual, and others asked to match aspects of the intervention (e.g. topic, duration) and record any teacher professional development that took place. In some studies, the science teaching of the control group was not observed.

Studies focused on a range of outcomes, with most focusing on attainment, but outcomes also reported on attitudes, critical thinking and argumentation and scientific language. A variety of research instruments and measures were used, some developed by the research team specifically for the intervention, whilst others used standardised instruments such as state or national tests. Examples included the Woodcock-Johnson-III language tests (Woodcock *et al.*, 2001), Measure of Academic Progress (MAP) and Primary Grade Reading (Northwest Evaluation Association, 2011), and DIBELS (Basic Early Literacy; Goffreda *et al.*, 2009), the California English Language Development Test, California State Test in English Language Arts, California State Test in Science, and WestEd (Klein & Bolus, 2006) and the Iowa Test of Basic Skills, for attainment in science and reading (see Vitale & Romance, 2012).

Findings

Studies using approaches based on language, literacy and text based approaches reported outcomes relating to attainment, attitudes, critical thinking and scientific language.

Table 17: Key features of studies on language, literacy and text based approaches in primary science teaching

Authors	Country	Design	Age of pupils	Sample pupils *schools*	Approach and assessment	Comparison	Outcomes measured	Summary of effects
Alexander (2018)	UK	RCT	9-10	4958 *76*	Dialogic teaching QA result: high	Teaching as usual	Attainment: standardised progress test in science	*Attainment* ES(R) =0.12
Bigozzi *et al.* (2011)	Italy	Quasi-experiment	8-11	172 *not stated*	Individual writing then group discussion then individual writing in context of observation during teacher demonstration QA result: low	Group discussion then individual writing in context of observation of teacher demonstration	Attainment: coding system for scientific conceptualization and metacognitive thinking applied to pupils' writing Critical thinking: written response to task	*Attainment* ES(R): *d*=0.34 to 1.20 (W1-D-W2 vs D-W); ES(R): *d*=0.02 to 1.10 (W1 vs D-W). *Critical thinking* data presented do not facilitate calculation of ES
Bravo & Cervetti, (2014)	USA	Quasi-experiment	9-11	172 *5*	Science and literacy integration: inquiry, reading, discussion, writing (space science) QA result: moderate	District-adopted space science unit	Attainment: researcher-constructed tests in science understanding Scientific language: researcher-constructed tests in science vocabulary, and reading comprehension.	*Attainment* data presented for whole group do not facilitate calculation of ES; ES(C): *g*=0.21 (for English Learners only) *Scientific language* ES(C): *g*=0.09 (whole group); ES(C): *g*=0.20 (ELs only)
Cervetti *et al.* (2012)	USA	RCT	9-10	2019 *not stated*	Science and literacy integration: reading, investigation, discussion, writing (light) QA result: high	Content-comparable science-only unit on light and energy using materials provided by school districts	Measures developed by the research and curriculum team coordinated by the Center for Research, Evaluation, and Assessment at the Lawrence Hall of Science Attainment: science knowledge test Scientific language:vocabulary and reading comprehension test	*Attainment* ES(C): *g*=0.45 *Scientific language* ES(C): *g*=0.26 (vocabulary) No effect for reading comprehension
Cheng *et al.* (2015)	Taiwan	Quasi-experiment	10-11	58 *1*	Adapted textbook with a cognitive principles-driven design (air and combustion) QA result:moderate	Standard textbook	Attainment: test of learning performance end-of unit examinations	*Attainment* ES(R): η^2=0.14
Connor *et al.* (2017)	USA	RCT	4-9	418 *6*	Content-area literacy instruction - science in the literacy block (various) QA result: high	Business as usual during the literacy block (no science)	Attainment: researcher-developed science knowledge test Scientific language: WJ-III Oral Comprehension test (Woodcock et al., 2001); researcher developed reading comprehension test	*Attainment* ES(C): *g*=1.50 *Scientific language* ES(C): *g*=0.05 (vocabulary)

Table 17 (continued)

Authors	**Country**	**Design**	**Age of pupils**	**Sample** pupils *schools*	**Approach and assessment**	**Comparison**	**Outcomes measured**	**Summary of effects**
Goldschmidt & Jung (2011)	USA	RCT	9-10	2019 *49*	Science and literacy integrated unit: reading, investigation, discussion, writing (light) QA result: high	present the content of their state science standards using regular curriculum materials on light	Attainment: science knowledge assessment Attitudes: student attitude assessment Scientific language: science assessment (reading and vocabulary)	*Attainment* ES(C): *g*=0.45 *Attitudes* data presented do not facilitate calculation of ES *Scientific language* ES(C): *g*=0.26 (vocabulary) No effect on reading
Hanley *et al.* (2015)	UK	RCT	9-10	1264 *42*	Thinking, Doing, Talking Science (various topics) QA result: high	Teaching as usual with waitlist	Attainment: test developed by evaluator using standardised assessment questions Attitudes: questionnaire adapted from Kind, Jones and Barmby (2007)	*Attainment* ES(R): *g*=0.22 (with larger effect for FSM pupils: ES(R): *g*=0.38) *Attitudes* data presented do not facilitate calculation of ES
Hanley *et al.* (2020)	UK	RCT	9-10	1264 *42*	Thinking, Doing, Talking Science (various topics) QA result: high	Teaching as usual with waitlist	Attainment: test developed by evaluator using standardised assessment questions Attitudes: questionnaire adapted from Kind, Jones and Barmby (2007)	*Attainment* ES(R): *g*=0.22 (with larger effect for FSM pupils: ES(R): *g*=0.38) *Attitudes* data presented do not facilitate calculation of ES
Henrichs & Lesemann (2014)	Netherlands	RCT	5-6	241 *31*	Teacher training in small group discussion integrating science and language learning (air pressure and reflection) QA result: moderate	Same tasks and topics as the experimental group teachers but without any training	Scientific language: Measures of incidences of on-task utterances and words, scientific reasoning and lexical diversity coded by researchers	*Scientific language* ES(R): *d*=0.63 (air pressure task only) No effect on reflection task
Jay *et al.* (2017)	UK	RCT	9-10	1233 *76*	Dialogic teaching QA result: high	Teaching as usual (not specified)	Attainment: standardised progress Test in Science	*Attainment* ES(R) =0.12
Kim *et al.* (2021)	USA	Cluster RCT	6-8	5494 *30*	Model of reading engagement (content literacy integration) in context of animals and habitats QA result: high		Critical thinking: argumentative writing assessed using a rubric designed by researchers Scientific language: semantic association task (Read, 1998, 2004) used to measure vocabulary depth, Measure of Academic Progress (MAP) Primary Grade Reading (Northwest Evaluation Association, 2011) used to measure comprehension	*Critical thinking* ES(R): *d*=0.24 *Scientific language* ES(R): *d*=0.50 (vocabulary) No significant effects on reading and basic literacy

Table 17 (continued)

Authors	Country	Design	Age of pupils	Sample pupils *schools*	Approach and assessment	Comparison	Outcomes measured	Summary of effects
Kitmitto *et al.* (2018)	UK	Cluster RCT	9-10	8996 *205*	Thinking, Doing, Talking Science (various topics) QA result:high	Business as usual waitlist	Attainment: test developed by evaluator using standardised assessment questions Attitudes: questionnaire adapted from Kind, Jones and Barmby (2007)	*Attainment* ES(R): *g*=0.01 (with larger effect for FSM eligible pupils: ES(R): *g*=0.05) *Attitudes* ES(R): *g*=0.12 (interest); ES(R): *g*=0.09 (self-efficacy)
Lai & Chan (2020)	Taiwan	Quasi-experiment	10-11	118 *1*	Integrating science with reading textbooks QA result: low	'Traditional instruction'	Attainment: researcher-developed science achievement test Attitudes: researcher-developed attitudes toward science scale	*Attainment* ES(C): *g*=0.23 *Attitudes* ES(C): *g*=0.11
Llosa *et al.* (2016)	USA	RCT	10-11	6673 *66*	Promoting Science for English Language Learners: inquiry and language development QA result: high	District adopted textbooks	Attainment: State science assessment used as outcome measure, researcher-developed science assessment as pre-measure	*Attainment* ES(R): *d*=0.25 (on researcher-developed test); ES(R): *d*=0.15 (on state test)
Maerten-Rivera *et al.* (2016)	USA	Cluster RCT	10-11	20879 *63*	Promoting Science for English Language Learners: inquiry and language development QA result: high	District-adopted science curriculum	Attainment: State science assessment	*Attainment* ES(R): *OR*=0.98 (year 1); ES(R): *OR*=1.28 (year 2); ES(R): *OR*=1.36 (year3)
Vitale & Romance (2012)	USA	Quasi-experiment	6-8	363 *2*	Science IDEAS curriculum based on cognitive science integrated science and reading QA result: high	Regular district science programme	Attainment: ITBS Science subtest Scientific language: ITBS Reading Comprehension subtest	*Attainment* data presented do not facilitate calculation of ES *Scientific language* data presented do not facilitate calculation of ES
Zwiep & Straits (2013)	USA	Quasi-experiment	4-11	3347 *5*	5E with English language development QA result: moderate	District English language development curriculum	Attainment: state mandated English and science assessments	*Attainment* ES(C): *g*=0.36 (for students who had the intervention for two years)

Attainment (knowledge and conceptual understanding)

A total of 16 studies reported impacts of language, literacy and text based approaches on attainment.

First, dialogic teaching and Thinking, Doing, Talking Science are discussed. Talk plays a key role in these studies. The second set of studies integrate science and literacy teaching. The third reported on approaches for teaching science to English language learners. Finally, studies using text-based approaches are discussed.

Two studies, drawing on the same dataset for the effect on attainment, focused on dialogic teaching (Alexander, 2018; Jay *et al.,* 2017). Dialogic teaching is described as a set of repertoires underpinned by psycho- and socio-linguistic, neuroscientific and philosophical principles which can be used to energise teacher and pupil talk. It is based on principles of collectivity, reciprocity, support, cumulation and purpose. The aim of such dialogue is to elicit pupils' everyday, common-sense perspectives, engage with their developing ideas and help them overcome misunderstandings. The intervention took place over two years and included a very substantial CPD component for teachers. The studies found a positive effect of dialogic teaching on attainment in science, reported as equivalent to pupils making on average two months additional progress when compared with pupils who did not receive the intervention.

Three studies related to Thinking, Doing, Talking Science, a professional development programme which aims to increase conceptual challenge by encouraging higher order thinking. An efficacy trial was reported by Hanley *et al.* (2015) and written up with the programme designers (Hanley *et al.*, 2020). An effectiveness trial was reported by Kitmitto *et al.* (2018). Mixed effects on attainment have been found, with Hanley *et al.* (2015; 2020) reporting positive effects, and a null effect reported by Kitmitto *et al.* (2018). Changes were made to the CPD model (fewer sessions, presented by trained trainers rather than the originators of the programme, and less financial support for schools) for the effectiveness trial, which may, as the evaluators also acknowledged, have had a negative influence on its effectiveness.

Five studies reported effects of integrating science and literacy teaching. Two large-scale studies (Cervetti *et al.*, 2012, and Goldschmidt & Jung, 2011) drawing on the same dataset used randomised control trial designs to investigate the integration of reading, speaking and writing into the teaching of science and found positive effects on science attainment. Cervetti *et al.* (2012) embedded cognitive strategy instruction into a curriculum which balances science and literacy instruction by, for example, selecting roles for texts in support of inquiry, modelling how scientists use texts during inquiry and using discussion to support sense-making. A moderate positive effect was found when compared with a control group learning the same topic with content matched to the state standards using a test which aligned with state standards and the content in the treatment unit. A large effect was calculated for the study by Goldschmidt and Jung (2011) which reported on the Seeds of Science/Roots of Reading project, which involved explicit instruction of literacy including comprehension strategies, making predictions, summarising, using non-fiction text structures to locate information and engaging in talk. These positive effects are supported by three smaller scale studies. One randomised control trial (Connor *et al.*, 2017) tested the effects of content-area literacy instruction (CALI) linking science, social science and written and oral comprehension. A large effect size was calculated, based on proximal content knowledge assessments, and with the note that students in the CALI group are likely to have received more science instruction than the control group. Also, a quasi-experimental study by Vitale and Romance (2012) investigated the impact of the Science IDEAS (In-depth Expanded Applications of Science) programme on basic science skills.

Science IDEAS is based on cognitive science and requires explicit representations of core concepts and relationships. Vitale and Romance reported a significant difference in favour of the Science IDEAS group. Finally, a relatively small study (Bigozzi *et al.*, 2011) investigated how short discussions before and/or after writing sessions impact scientific language development. They found that asking pupils to reflect individually on a practical demonstration they have just witnessed, and write notes on their observations, before group discussion on these observations, is more effective than discussion without the prior reflection.

Four studies reported specifically on programmes to support English language learners. Both Llosa *et al.* (2016) and Maerten-Rivera *et al.* (2016) reported on the use of Promoting Science Among English Language Learners (P-SELL) which integrates an inquiry approach to science teaching with teacher CPD and school science resources. P-SELL aims to improve science achievement of all students, but especially English language learners. They found positive impacts on science conceptual understanding (Llosa *et al.*, 2016; Maerten-Rivera *et al.*, 2016) using both standard state-level science assessments and a researcher-developed science test. Llosa *et al.* (2016) also reported positive effects on attainment in English. However, Maerten-Rivera *et al.* (2016) found that it is not until the third year of a programme like P-SELL, with teachers receiving CPD throughout the programme (mostly in the summer holidays but also throughout the year), that significant changes are perceived in pupil outcomes as measured by standardised assessments. Bravo and Cervetti (2014) reported a small positive effect on attainment for English language learners as a result of a quasi-experimental study of a curriculum intervention which incorporated first-hand inquiry with reading and discussion in support of inquiry, and writing about investigations and science concepts. Finally, Zwiep and Straits (2013) blended scientific inquiry using the 5E instructional model with pupils' English language development, finding small positive effects on both science and language tests, although the control group condition was the district's English Language Development programme, alongside "very little, if any, instruction in science" (p1317).

Two studies focused on the use of textbooks. Lai and Chan (2020) found a small positive effect of integrating science trade book reading strategies on pupils' science understanding. Reading strategies were broadly defined and included reading the text, conducting experiments, drawing mind maps and participating in discussions. Another small-scale study of textbooks by Cheng *et al.* (2015) found that considered but express use of illustrations in textbooks can significantly enhance pupils' understanding of the concept or skills described in the text, while misuse of illustrations can lead to considerable confusion and distraction.

Attitudes, interest and motivation

Five studies reported on attitudinal outcomes: three on Thinking, Doing, Talking Science, one on the use of textbooks, and one on science and literacy integrated instruction. The most recent evaluation of Thinking, Doing, Talking Science reports small positive effects on interest and self-efficacy (Kitmitto *et al.*, 2018). Effect sizes were not calculated for the early study on Thinking, Doing, Talking Science (Hanley *et al.*, 2015), nor for the study on science and literacy integrated instruction (Goldschmidt & Jung, 2011). In contrast, Lai and Chan (2020) found no meaningful effect of integrating science trade book reading strategies on pupils' attitudes towards science.

Critical thinking and argumentation

Two studies reported outcomes related to critical thinking. Kim *et al.* (2021) conducted a cluster randomised control trial on a content literacy intervention, the Model of Reading Engagement (MORE) which consisted of thematic lessons, concept mapping and interactive read-alouds. They reported a small positive effect on argumentative writing in science, with a greater effect for initially higher performing pupils than their peers, with vocabulary knowledge depth mediating the effects of the intervention. Bigozzi *et al.* (2011) investigated the effect of individual writing on metacognitive thinking, assessed by measuring pupils' awareness about the distinction between appearance and reality, and changing their ideas. Conclusions are less clear as the data to assess these outcomes comes from different tasks relating to different experiments in different school grades.

Scientific language and communication

Seven studies reported scientific language outcomes, all of which focused on science and language or literacy integration.

Positive effects were reported on all studies, except for the study of a cognitive science-based instructional model integrating science, reading and writing (Science IDEAS) by Vitale and Romance (2012). Whilst no effect size was reported, the experimental group was found to perform significantly better in science and reading. Cervetti *et al.* (2012) and Goldschmidt and Jung (2011) used approaches partly inspired by Science IDEAS, both reporting a positive effect on writing and vocabulary. Other related approaches to science and language integration included content-area literacy instruction (CALI) (Connor *et al.*, 2017), which was developed using the 5E learning cycle (Bybee, 1997), and which influenced the Model of Reading Engagement (MORE) approach (Kim *et al.*, 2021). Studies of both the CALI and MORE approaches reported positive effects of the intervention on scientific language. Henrichs and Lesemann (2014) integrated science and language through discussion with young (age 5) children. The discussions followed professional development focused on academic language input and how to prompt children to elicit academic language. They found a moderate effect on lexical diversity, including the use of domain-specific words in the air pressure task, but not for the mirrors task.

Finally, in the context of working with English language learners in science, Bravo and Cervetti (2014) investigated the effect of a curriculum development in which language processes (e.g. reading, writing) are used to teach science. They found a small positive effect on vocabulary but not on reading when compared to a group with a focus on 'hands-on' science.

Implementation and resources

Barriers to implementation of longer interventions were reported. These included teacher time, curriculum pressures, staff and pupil turnover and unexpected events such as illness. All of the studies were conducted in regular classrooms, taught by the regular classroom teacher. Some of the programmes required consistent and considerable commitment, for example, the implementation of a science curriculum in its entirety, or to several instances of CPD. Some of the approaches required teachers to undertake CPD during the summer vacation period, and in some cases it took several years for the approach to embed fully into a teacher's practice before the impact was seen on pupil outcomes.

Costs were not usually reported, apart from those related to Thinking, Doing, Talking Science (£29/pupil/year) and dialogic teaching (£52/pupil/year).

Some studies involved the use of resources or curriculum units, with some materials (e.g. Henrichs & Leseman, 2014) no longer available online.

Opportunity cost was highlighted as overwhelmingly in favour of science/literacy integration as progress in English was just as good as for separate structured English language development programmes, while time was taken from literacy lesson time.

Gaps in the evidence base

No studies were found to focus on the role of supplementary support such as teaching assistants in relation to language and literacy in a science context. The theme of language development for EAL pupils in particular, would benefit from an understanding of the role of supplementary support.

Whilst several studies in this cluster had an emphasis on the development of science in pupils whose first language is not English, either on its own or in tandem with enhancing their English proficiency, there is little evidence of the benefits or otherwise of using diverse languages in the classroom during science instruction.

The majority of studies focused on attainment in Science or English as an outcome, with fewer focusing on other outcomes - and none on outcomes relating to working scientifically and problem solving.

This cluster included the full age spectrum of children in primary schools in England, although fewer focused on younger children.

Quality of the evidence base

This cluster had several high-quality studies, and several studies focusing on the same or related approaches. All studies on extended interventions such as P-SELL and Thinking, Doing, Talking Science were judged as high-quality. The randomised control trials were of appropriate size, as guided by power calculations. Findings were fairly consistent (most reported positive or null effects across a range of outcomes). Reasons for low-quality ratings included no reporting of differences between intervention and control groups.

Unusually for studies reported in this review (and more widely for studies reporting on educational interventions), an external evaluation of several interventions was undertaken in addition to research conducted by the research team (e.g. Hanley *et al.*, 2015; Kitmitto *et al.*, 2018; Jay *et al.*, 2017)

Science/literacy integration approaches were fairly consistent, blending science and English in order to develop (academic) language as well as science conceptual understanding, with extensive curriculum materials and CPD for teachers.

Applicability of the evidence base

The studies focusing on talk in science (dialogic teaching, Thinking, Doing, Talking Science) have been conducted in England with large numbers of pupils.

Applicability of the approaches for teaching science to English language learners is likely to vary depending on the linguistic diversity of pupils in schools. While some adaptations may be necessary, the approach to integrating science with literacy is broadly applicable to the English context. The studies conducted in the USA tended to use designs which integrate innovations into a curriculum which values inquiry and scientific practices as well as scientific content. This might be a key difference with baseline approaches used in schools in England.

Successful implementation of a science and literacy integrated curriculum is likely to require substantial CPD for teachers. However, a recent survey for Wellcome (Leonardi *et al.,* 2017 found that 30% of teachers reported that they “had not received any support for science teaching in the last year” from their school, and that only 37% of schools had an allocated budget for CPD in science. The high stakes assessment system has been found to narrow the curriculum, with primary schools placing greater focus on literacy and numeracy than other subjects (Hutchings, 2015; UK Parliament, 2017). Indeed, in the survey by Leonardi *et al.* (2017), only 30% teachers rated science as very important, compared to 80% for English and maths. The evidence points to the success of science specific CPD, in increased enthusiasm for science in both teachers and their pupils, alongside academic achievement.

Overall evidence statement/Key findings

Multiple large-scale, high-quality studies point to positive (or at worst, null) impacts on attainment and scientific language of integrating science and literacy. The evidence also suggests that approaches that are designed to improve outcomes for pupils who are less proficient in English result in differentially positive outcomes in science for such pupils, without compromising progress for all pupils. In addition, such approaches result in improved attitudes towards science, especially for English language learners because of the stimulation from content other than English language curricula.

In terms of attitudinal outcomes, there were positive and null results for the small number of approaches focused on talk and reading in science. Similarly, there was a positive effect on critical thinking, but only two studies reported on these outcomes.

Table 18: Summary of the evidence on language, literacy and text-based approaches

Number of studies	18
Assessed quality of studies	High: 12 Moderate: 4 Low: 2
Location of studies	Italy: 1 Netherlands: 1 Taiwan: 2 UK: 5 USA: 9
Outcomes of focus	Attainment (knowledge and conceptual understanding): 16 Attitudes, motivation and interest: 5 Critical thinking and argumentation: 2 Scientific language: 7
Design	Quasi-experimental design: 6 Randomised control trial: 12
Consistency of findings	Most language and literacy interventions reported a positive (or null) impact on science attainment and language.
Summary of the evidence base	• Study design: multiple randomised control trials • Multiple high-quality studies with large sample sizes • No important inconsistency • Several studies conducted in a representative population in England • Standardised instruments used • Although many interventions were reported/investigated by the developers of the intervention itself, there were some instances of independent evaluators

B8 Learning outside the classroom

Definitions

In this review, *learning outside the classroom (LOC)* is characterised by whole class activities managed by the class teacher away from the classroom. This includes learning at museums, science centres, planetaria, at industrial sites, through science trails, at zoos, farms, botanic gardens, residential centres and on school grounds (Braund & Reiss, 2004).

Description of the evidence base

The learning outside the classroom cluster consists of six studies, presented in Table 19.

The studies in this cluster presented a variety of approaches to learning outside the classroom, from fieldwork and the effect of activities supporting a field trip (Glick & Samarapungavan, 2008; Mills & Katzman, 2015; Scott & Boyd, 2016; Wünschmann *et al.* 2017), school gardens (Wells *et al.*, 2015), and a STEAM (Science, Technology, Engineering, Arts and Mathematics) intervention (Piila *et al.*2021).

The studies were conducted mainly in biological context, with one taking place at a gravitational wave observatory and another in the context of a Mars-colonisation module. Half of the studies were undertaken in the USA, with the others in Finland, Germany and the UK. The studies involved children aged 5-12, although only one reported on the approach with children under the age of eight.

Four studies adopted a quasi-experimental design, with one a randomised control trial and the other a randomised field trial. One study was assessed to be high-quality, three moderate and two low-quality using the quality assessment instrument. In this cluster, the control groups were more clearly described with only two providing limited details of teaching as usual.

Outcomes of focus included attainment and attitudes. A range of measures were used, including researcher-designed tests and scoring systems for interviews on the topic (Glick & Samarapungavan, 2008; Wells *et al.*, 2015), items from standardised tests (Scott & Boyd, 2016), a science knowledge test not specific to topic (Piila *et al.*, 2021), validated scales such as Possible Science Selves (Mills & Katzman, 2015) and an adaptation of the intrinsic motivation inventory (Wünschmann *et al*., 2017).

Findings

Studies using approaches based on learning outside the classroom reported outcomes relating to attainment and attitudes. Where effect sizes were provided or calculated, these tended to be positive (or null) in favour of learning outside the classroom. There was little consistency in approach or outcome measure and we advise against comparing effect sizes.

Table 19: Key features of studies on learning outside the classroom in primary science teaching

Authors	Country	Design	Age of pupils	Sample pupils *schools*	Approach and assessment	Comparison	Outcomes measured	Summary of effects
Glick & Samarapungavan (2008)	USA	Quasi-experiment	9-10	30 *2*	School trip with use of field notebook and activities before and after the school trip (animals) QA result: low	School trip without supporting activities	Attainment: pre- trip knowledge test, post- trip scored interview	*Attainment* ES(R): η^2=0.25
Mills & Katzman (2015)	USA	Quasi-experiment	10-11	151 *1*	School trip to science research and learning centre including meeting and interviewing scientists QA result: moderate	School trip to science research and learning centre meeting educator instead of scientists	Attitudes: Possible Science Selves (Beier et al., 2012) including measures of desire and participation	*Attitudes* ES(R): *d*=0.40 (PSS desire) No effect on participation
Piila *et al.* (2021)	Finland	Quasi-experiment	10-12	364 *8*	STEAM-based learning of Mars colonisation QA result: low	STEM subjects taught in 'a more traditional way' (no further details)	Attainment: science knowledge test (not specific to topic)	*Attainment* data presented do not facilitate calculation of ES
Scott & Boyd (2016)	UK	Quasi-experiment	9-11	379 *8*	Fieldwork in local habitat to identify, photograph and produce field guide QA result: moderate	No fieldwork, taught the ecological content in class	Attainment: test using previous science SAT questions, written assessments post-intervention	*Attainment* data presented do not facilitate calculation of ES
Wells *et al.* (2015)	USA	RCT	5-12	3061 *49*	School garden and curriculum intervention: raised beds, lesson plans and other resources QA result: high	No garden (waitlist) control group, not specified what they did	Attainment: multiple-choice questionnaire from the University of Missouri (UM) 'Eating from the Garden Curriculum' survey	*Attainment* ES(C): *g*=5.86
Wünschmann *et al.* (2017)	Germany	Randomised field trial	8-10	65 *1*	Visiting amphibians and reptiles in a zoo QA result: moderate	Group taught about reptiles and amphibians in school and group without teaching on the topic	Attainment: researcher designed test of reptile knowledge Attitudes: German adaptation (Wilde *et al.,* 2009) of the intrinsic motivation inventory (Deci & Ryan 2003)	*Attainment* ES(R): η^2= 0.41 *Attitudes* ES(R): η^2=0.12 (perceived choice); no significant difference for interest/enjoyment

Attainment (knowledge and conceptual understanding)

Five studies in the cluster explored the effect of learning outside the classroom on outcomes related to attainment. These consisted of two long-term programmes (Wells et al., 2015; Piila *et al.*, 2021) and three briefer interventions (Glick & Samarapungavan, 2008; Scott & Boyd, 2016; Wünschmann *et al.*, 2016).

The long-term programmes included the Healthy Gardens, Healthy Youth initiative (Wells *et al.*, 2015) to support teaching plant science and nutrition in low-income schools. Teachers were provided with an educational toolkit, raised-bed and container garden kits, a garden implementation guide, and extension educators to facilitate weekly activities in the garden for the duration of a year. The classes in the control group continued covering the same curriculum material in standard classroom sessions. The randomised control trial found a significant modest positive effect of the garden intervention on nutrition and plant knowledge, although it is not clear what the control group had been taught about these topics, and the amount of time spent on the garden lessons mattered. Another long-term learning outside the classroom intervention was study of the cross-curricular module on Mars colonisation reported by Piila *et al.* (2021). The intervention involved classroom visits to a planetarium and interactive science centre. Teachers were provided with lesson outlines and ideas for activities. In addition to the out-of-school learning aspect, the approach made use of scientific technology such as augmented reality and international exchange platforms. This multi-pronged approach took several months. The classes in the control group were taught the same topic but through a different theme. Piila *et al.* (2021) reported improved achievement for pupils who participated in the intervention, although effect sizes were not reported.

The three other studies reported the impact of brief out-of-school experiences: a half-day ecological fieldwork activity in a hedgerow, playground or park (Scott & Boyd, 2016), a visit to a reptile and amphibian zoo (Wünschmann *et al.*, 2016) and a trip to a wolf park (Glick & Samarapungavan, 2008). Scott and Boyd (2014) reported a significant effect on science achievement (effect size not calculated) of the field trip. Similarly, Wünschmann *et al.*, (2016) found positive effects of a field trip to a reptile and amphibian zoo, when compared to classroom teaching on the topic and to classroom teaching on a different topic. Glick and Samarapungavan (2008) found a positive effect of participation in the researcher-designed field trip-related classroom activities before and after the field trip on pupil learning.

In terms of effects on specific groups of pupils, Piila *et al.* (2021) reported significantly increased learning outcomes for high-achieving pupils, with no significant improvement for low achievers.

In conclusion, these five studies together indicate that learning outside the classroom (of a short or longer duration) can have a positive effect on science achievement of pupils of primary school age.

Attitudes, interest and motivation

Two studies focused on attitudinal and related outcomes from learning outside the classroom, both based on half day visits. The study by Wünschmann *et al.* (2016) found no effect on three components of motivation (e.g. interest/enjoyment, perceived competence, and pressure/tension) but a small positive effect for pupils in the two visit groups on perceived choice. Mills and Katzman (2015) found a positive effect of interviewing scientists during a field trip on desire to become a scientist. There was no effect on willingness to participate in science activities between the pupils who interviewed the scientists and those who did not.

The studies in the review are too few and too diverse to draw any conclusions about the effectiveness of different approaches to learning outside the classroom on improving pupils' attitudes.

Implementation and resources

The time taken to complete risk assessments, liaise with external partners, obtain approvals and organise visits are likely to be barriers to implementation. Wells *et al.* (2015) observed that teachers perceived garden-based lessons as an 'extra', even when aligned with the standards required of teachers.

Costs were not reported, however there is often a cost associated with visits beyond the school (e.g. for transport and entry to sites), and with the establishment of school gardens. There is also an additional demand of time on teachers to organise visits and maintain school gardens.

Gaps in the evidence base

A number of approaches to learning outside the classroom were absent from the studies included in the review. These include forest schools, science fairs, STEM (and space and climate) ambassador programmes and CREST (Creativity in Science and Technology) Awards (which have been evaluated with older pupils). These approaches were mentioned in the review of practice, and research on their effectiveness is likely to be of interest to teachers.

Another gap in the evidence base relates to outcomes. Whilst pupil attainment, and to a lesser extent, attitudes, tended to be measured in studies in this cluster, more relevant outcomes might relate to working scientifically.

Only one study involved children younger than eight years old. Given the popularity of outdoor learning programmes such as forest schools in Key Stage 1, a focus on younger children is likely to be of interest.

Quality of the evidence base

Using the quality assessment tool, one study was rated high, three were rated moderate and two were rated low.

Applicability of the evidence base

The studies described field trips, field work, visits to science centres and school gardens, all of which are applicable to schools in England. The high-quality large-scale RCT by Wells *et al.* (2015) which focuses on the use of school grounds is likely to be particularly interesting to teachers in England given the recent establishment of the National Education Nature Park (Natural History Museum, 2023), which aims to drive and increase engagement with nature for all children.

The applicability of Learning Outside the Classroom approaches depend on access to gardens and raised beds in primary schools and geographical proximity to other settings for primary science such as science centres and planetaria.

Overall evidence statement/Key findings

Evidence suggests that learning outside the classroom can have a positive effect on outcomes related to attainment (typically scientific knowledge). The evidence is less clear in relation to attitudinal outcomes such as motivation, desire and participation in science. The overall evidence base for this cluster is fair because the evidence is sufficient to determine effects on outcomes, but limited. Whilst the approaches are fairly generalisable to routine practice, the cluster consists of a small number of studies, most of which are fairly small-scale.

Table 20: Summary of the evidence on learning outside the classroom

Number of studies	6
Assessed quality of studies	High: 1 Moderate: 3 Low: 2
Location of studies	Finland: 1 Germany: 1 UK: 1 USA: 3
Outcomes of focus	Attainment (knowledge and conceptual understanding): 5 Attitudes, motivation and interest: 2
Design	Quasi-experiment: 4 Randomised control trial: 1 Randomised field trial: 1
Consistency of findings	Results were positive or null when compared with business-as-usual
Summary of the evidence base	● Study design: Two randomised trial design ● One large-scale high-quality study ● Consistent findings ● Variability in interventions and outcome measures ● All interventions except two were developed and evaluated by the researchers reporting the work ● All studies used different interventions and different instruments to assess outcomes

B9 Practical work, inquiry and investigation

Definitions

For the purposes of this review an *inquiry-based approach* is defined as a strategy using tasks where pupils (i) formulate questions or hypotheses, (ii) design one or more experiments to address these questions, (iii) execute the experiments, (iv) interpret the data they find, (v) draw conclusions and generate explanations, and (vi) communicate their findings. The level of teacher support and direction during this inquiry process may make the inquiry range between a guided and open inquiry. In contrast, a *practical work* approach describes "activities in which the students manipulate and observe real objects and materials" (Abrahams & Millar, 2008, p. 1945). These studies are united by their focus on 'doing' science. We describe the evidence base then present the findings on practical work and inquiry and investigation separately.

Description of the evidence base

The practical work, inquiry and investigation cluster consists of 18 studies, presented in Tables 20 (practical work) and 21 (inquiry and investigation).

The 18 studies in this cluster included 12 which reported the effect of an inquiry-based teaching approach and six which reported on practical work. A variety of approaches were found within each sub-cluster. Practical work included the use of practical stations, hands-on structured experiments, looking after animals and designing a prototype. Topics spanned biological, chemical and physical sciences. The inquiry and investigation sub-cluster included approaches such as argument-driven inquiry (with links to the critical thinking and argumentation cluster), the 5E instructional model, self-regulated inquiry, guided inquiry, design-based inquiry and inquiry with open problems.

The practical work sub-cluster consisted of small-scale single- or two-school studies conducted in Germany , Serbia, Switzerland and the USA, with two conducted in Turkey. The inquiry and investigation sub-cluster consisted of one large-scale and one small-scale study conducted in the USA, with a number of smaller scale studies in Argentina (1), China (1), Croatia (2), Switzerland (1) and five smaller studies carried out in single schools in Taiwan. No studies were conducted in the UK. Studies were conducted with the full range of primary-aged children (ages 4-11), although most focused on older primary school age, and only two, both in the practical work sub-cluster, worked with 5- 8 year-old pupils.

Most studies (16) adopted a quasi-experimental design, with two using randomised control trial (RCT) designs (Polikoff *et al.*, 2018; Zhang, 2018). One study was assessed to be high-quality, ten moderate-quality and seven low-quality. In the practical work sub-cluster, four studies were assessed to be moderate-quality and two low-quality. In the inquiry and investigation sub-cluster, one was assessed to be high-, six moderate- and five low-quality.

In the practical work sub-cluster, four studies focused on attainment as an outcome, with two reporting on critical thinking, and one each reporting on attitudinal outcomes and working scientifically. Measures tended to be developed by the research team, except for the motivation scale used by Meyer *et al.* (2016). In the inquiry sub-cluster, six studies reported on outcomes related to attainment, six on attitudes, and five each on working scientifically and critical thinking. Measures included tests developed by the research team, and also those used in earlier studies. For example, DiMauro and Furman (2016) used assessments validated by Furman (2012) and Lai *et al.* (2018) assessed information-seeking strategies, self-efficacy and self-regulation based on previous research (Chang & Tsai, 2011, Wang & Hwang, 2012, and Barnard *et al.* 2009).

Table 20: Key features of studies on practical work in primary science teaching

Authors	Country	Design	Age of pupils	Sample pupils *schools*	Approach and assessment	Comparison	Outcomes measured	Summary of effects
Cvjetićanin *et al.* (2015)	Serbia	Quasi-experiment	9-10	136 *not stated*	Practical work: hands on experience of structured experiments (physical and chemical properties of materials) QA result: low	Teacher demonstration of the same experiments.	Attainment: Tests including tasks to assess different levels of knowledge	*Attainment* ES(C): g=0.70
Dankenbring & Capobianco (2016)	USA	Quasi-experiment	10-11	67 *1*	Practical work: engineering design based science lessons - design a prototype (sun-Earth system) QA result: moderate	Traditional science lessons - book, graphic organiser, video, writing modelling task	Attainment: multiple choice topic knowledge test; draw and explain task	*Attainment* no significant effect; draw and explain data presented do not facilitate calculation of ES
Durmuş & Bayraktar (2010)	Turkey	Quasi-experiment	9-11	104 *1*	Practical work: laboratory experiments on matter and change QA result: moderate	Conceptual change text (matter and change)	Attainment: matter concept test developed by researchers	*Attainment* data presented do not facilitate calculation of ES
Leuchter *et al.* (2014)	Switzerland	Quasi-experiment	4-9	288 *not stated*	Practical work: learning activities with practical work stations (floating and sinking) QA result: moderate	Usual curriculum (no teaching on floating and sinking)	Attainment: researcher designed test of conceptual knowledge Critical thinking: response to reasoning task	*Attainment* ES(R): η^2=0.13 (at delayed post-test) *Critical thinking* ES(R): η^2=0.09 (at delayed post-test)
Meyer *et al.* (2016)	Germany	Quasi-experiment	10-11	166 *2*	Short and long-term contact with living animals (harvest mice) QA result: moderate	No living animals	Attitudes: intrinsic Motivation Inventory (Ryan 1982), Flow Short Scale (Rheinberg *et al.,* 2003)	*Attitudes* ES(R): η^2=0.05 to 0.09 (motivation); ES(R): η^2=0.29 (flow experience)
Ünal & Aral (2014)	Turkey	Quasi-experiment	5-6	42 *2*	Experiment Based Education Program QA result: low	Routine educational programme	Working scientifically: Problem Solving Scale in Science Education developed by researchers	*Working scientifically* ES(C): g=3.63

Table 21: Key features of studies on inquiry and investigation in primary science teaching

Authors	**Country**	**Design**	**Age of pupils**	**Sample** pupils *schools*	**Approach and assessment**	**Comparison**	**Outcomes measured**	**Summary of effects**
Chen & She (2015)	Taiwan	Quasi-experiment	9-10	116 *1*	Scientific inquiry with reasoning component QA result:moderate	Scientific inquiry without reasoning component	Attainment: Scientific Concepts Test Critical thinking: Scientific Concept-Dependent Reasoning Test Working scientifically: Scientific Concept-Dependent Inquiry Test	*Attainment* ES(R): η^2=0.13 *Critical thinking* ES(R): η^2=0.17 *Working scientifically* ES(R): η^2=0.2
Chen *et al.* (2019)	Taiwan	Quasi-experiment	9-10	68 *2*	Modified Argument-Driven Inquiry (MADI) QA result: moderate	Regular science lessons with structures hands-on work	Attitudes: Engagement in Learning Science Questionnaire Critical thinking: Argumentation ability tests	*Attitudes* ES(R): η^2=0.16 *Critical thinking* ES(R): η^2=0.12
Di Mauro & Furman (2016)	Argentina	Quasi-experiment	9-10	60 *1*	Inquiry unit with open problems QA result: moderate	Traditional instruction, mainly with texts and questionnaires	Working scientifically: test based on province learning goals and previously validated assessments e.g. Furman (2012)	*Working scientifically* data presented do not facilitate calculation of ES
Lai (2016)	Taiwan	Quasi-experiment	8	106 *1*	invitation - Prediction-operation-discussion (iPod inquiry) QA result: low		Attainment: Air Concept Comprehension Test Attitudes: Scientific Attitude Scale	*Attainment* ES(C): *g*=1.02 *Attitudes* ES(C): *g*=0.24
Lai *et al.*(2018)	Taiwan	Quasi-experiment	10	56 *1*	Self-regulated science inquiry QA result: moderate	Conventional science inquiry (same activities as treatment group)	Attitudes: questionnaires of self-efficacy (based on Wang & Hwang, 2012) and self-regulation (based on Barnard *et al.*, 2009) Working scientifically: Questionnaire on information seeking strategies, developed by teachers based on Cheng and Tsai (2011)	*Attitudes* ES(R): η^2=0.23 (learning approach); ES(R): η^2=0.87 (self-regulation); ES(R): η^2=0.11 (overall combined effect) *Working scientifically* ES(R): η^2=0.10 (information seeking strategies)
Li *et al.* (2016)	China	Quasi-experiment	10	30 *1*	Design-based engineering methodology QA result: low		Critical thinking: Problem-Solving Ability Self-Checking Questionnaire (Li, 2003)	*Critical thinking* ES(R)=0.40 (problem-solving ability)

Table 21 (continued)

Authors	**Country**	**Design**	**Age of pupils**	**Sample** pupils *schools*	**Approach and assessment**	**Comparison**	**Outcomes measured**	**Summary of effects**
Lin *et al.* (2009)	Taiwan	Quasi-experiment	10-11	92 *1*	Inquiry activities QA result: moderate	Teaching as usual, hands-on opportunities with instructions	Attitudes: Learning environment question based on WIHIC (Aldridge et al., 1999) Critical thinking: Classification of Student Questioning Skills Working scientifically: Scoring of Inquiry Activities	*Attitudes, Critical thinking, Working scientifically* data presented to not facilitate calculation of ES
Letina (2016)	Croatia	Quasi-experiment	9-10	333 *8*	Inquiry-based science (life conditions) QA result: low	'Traditional' science lessons	Attainment: researcher-designed test of scientific competence (knowledge and skills for solving problems and their capability for argumentative reasoning)	*Attainment* ES(C): *g*=2.23
Letina (2020)	Croatia	Quasi-experiment	9-10	333 *8*	Inquiry-based teaching (life conditions) QA result: low	Traditional instruction in the same curriculum content.	Attitudes: learning to learn competence questionnaire adapted from a questionnaire by Institute for Social Research in Zagreb	*Attitudes* ES(C): *g*=0.38
Polikoff *et al.* (2018)	USA	Cluster RCT	9-10	1651 *17*	5E (Engage, Explore, Explain, Elaborate, Evaluate) (speedometry) QA result: high	'Business as usual' with waitlist control (not speedometry)	Attainment: Content Knowledge Assessment developed with independent research institution Attitudes: Interest and Emotions Survey (Danielson *et al.*, 2016).	*Attainment* ES(R)=0.48 standard deviations *Attitudes*: no significant effect on interest

Table 21 (continued)

Authors	Country	Design	Age of pupils	Sample pupils *schools*	Approach and assessment	Comparison	Outcomes measured	Summary of effects
Schalk *et al.* (2019)	Switzerland	Quasi-experiment	8-11	189 *not stated*	Guided inquiry (physics) QA result: moderate	Traditional instruction (not physics) with waitlist	Attainment: Content knowledge test Working scientifically: Control of Variables test	*Attainment* ES(R): R^2=0.12 (floating and sinking); ES(R): R^2=0.76 (air and atmospheric pressure); ES(R): R^2=0.40 (sound); ES(R): R^2=0.66 (stability of bridges) *Working scientifically* data presented do not facilitate calculation of ES
Zhang (2018)	USA	Not specified (cluster RCT)	9-11	136 *2*	Hands-on investigation and withholding answers (physics) QA result: low	'hands-on' + no 'withholding' and no 'hands-on' + no 'withholding' (direct instruction) (light)	Attainment: content test on light Critical thinking: reasoning with learned content	*Attainment* no significant effect *Critical thinking* ES(C): 0.54-0.57 for hands-on, with stronger effect when answers not withheld.

Findings

Studies using approaches based on practical work, inquiry or investigation reported outcomes relating to attainment, attitudes, critical thinking and working scientifically and problem solving. There was little consistency in approach or outcome measure so we advise against comparing effect sizes of different approaches.

Attainment (knowledge and conceptual understanding)

Impacts of practical work, inquiry and investigation on attainment or conceptual understanding were reported in eleven studies.

Four small-scale quasi-experimental studies investigated the effect of practical work on outcomes related to attainment and conceptual understanding. Two of these reported positive results. Durmuş and Bayractar (2010) investigated the optimal way of addressing pupils' misconceptions, comparing conceptual change texts and structured hands-on experiments using a Predict-Observe-Explain sequence. Both treatment groups were reportedly more successful than the control group in overcoming misconceptions in the post-test and the delayed post-test, although neither significance nor effect sizes were provided. Leuchter *et al.* (2014) investigated the effect of practical work on conceptual understanding. They created a structured and problem-based learning environment with four work stations including hands-on activities organised according to a presumed conceptual progression. Although they found statistically significant gains in conceptual understanding, with a small effect size, this is not surprising as the control group was taught a different topic. Two studies reported null effects. Dankenbring and Capobianco (2016) found no significant differences between pupils in a group using practical work with an engineering design emphasis which required pupils to design a prototype structure to shade a picnic table compared with a control group which experienced teaching with a trade book and graphic organiser, video and reflective writing and a modelling exercise and Cvjetićanin *et al.* (2015) found no difference in remembering, understanding and application when they compared teacher demonstrations with hands-on practical experiences. However, they did find differences in favour of the hands-on experiment group in analysis, evaluation and creation. Whilst these studies provide very limited evidence to support the claim that practical work is effective in improving science conceptual understanding, nor is there evidence to suggest it is disadvantageous to conceptual understanding.

Six studies of inquiry or investigative approaches reported on outcomes related to attainment. Reports on the impact of inquiry on attainment come from two RCTs in the USA (Polikoff *et al.*, 2018; Zhang, 2018), two small quasi-experimental studies in Taiwan (Cheng & She, 2015; Lai *et al.*, 2018), one in Switzerland (Schalk *et al.*, 2019) and one in China (Li *et al.* (2016). There was considerable diversity in approaches used. Two used toys – Hot Wheels© (Polikoff *et al.*, 2018) and LEGO© (Li *et al.*, 2016) – and one (Lai *et al.*, 2018) used computer software in addition to materials usually found in science classrooms. Two (Polikoff *et al.*, 2018; Lai *et al.*, 2018) had in common the use of the 5E (Engage Explore Explain Elaborate Evaluate) instructional model, and Lai (2016) used iPod (inquiry-prediction-operation-discussion) inquiry. Chen and She (2015) used driving questions with open inquiry and Li *et al.* (2016) used an engineering design approach which added analysis, design and optimisation steps. Schalk *et al.* (2019) and Zhang (2018) used a guided inquiry approach. Schalk *et al.* (2019) designed content-focused physics curriculum units which included the control of variables in investigations, without explicit teaching of the strategy. Zhang (2018) investigated the effect of withholding answers. Letina (2016) compares an inquiry based teaching strategy which uses scientific methods and procedures e.g. observation, hypothesis creation and experimentation to lecture-based teaching.

The results from these studies are mixed. Li *et al.* (2016) found no statistically significant difference between the control group and experimental group, suggesting no impact of the modified engineering design approach on attainment of scientific knowledge. This is consistent with the study by Dankenbring and Capobiano (2016) of the effect of a practical work approach with an engineering design focus. Zhang (2018) found no effect when hands-on investigation was compared to direct instruction and to a hands-on condition in which answers were withheld.

In contrast, five studies reported positive impacts on learning. Schalk *et al.* (2019) reported small to large effects of guided inquiry on domain specific content knowledge across the units. Similarly, gains in conceptual knowledge were found for pupils participating in interventions required to make their reasoning explicit during inquiry work (Chen & She, 2015) and for pupils who took part in a Speedometry 5E curriculum intervention (Polikoff *et al.*, 2018). Lai (2016) found a positive effect of their approach which integrated science with life experiences, hands-on participation and discussion. Furthermore, statistically significant gains in learning achievement, information seeking tendency and formal query tendency were reported for experimental groups using open inquiry with self-regulated guidance (Lai *et al.*, 2018). Letina (2016) found positive effects of inquiry compared with lecturing on scientific competence. This could link to other outcomes but as few details of the questions are provided, this is not clear.

Whilst mixed, this set of studies together indicate that practical work and inquiry - particularly when guided or when there are prompts for reasoning – can have a positive effect, or a null effect, on primary school pupils' attainment in science.

Attitudes, interest and motivation

Six studies focused on attitudinal and related outcomes from practical work, inquiry and investigation.

One study focused on the impact of practical work on attitudinal outcomes. Meyer *et al.* (2016) reported on the effect of involvement in mouse keeping in the classroom on pupils' attitude to science. They found positive effects on interest and flow experience amongst pupils in classrooms with mice, particularly those with mice over a longer period, than in the classrooms with videos on laptops. Whilst this suggests a role for animals in education, it is of course important to consider animal welfare (social needs, compatibility with the school day and housing), potential allergic reactions amongst children, as well as staffing responsibilities. CLEAPSS (n.d.) provides advice on keeping animals in schools.

Five studies explored the effect of an inquiry approach on pupils' attitudes to science. The guided inquiry-based speedometry intervention (Polikoff *et al.*, 2018) and Modified Argument-Driven Inquiry approach (Chen *et al.*, 2018) have been discussed previously. The approach used by Lin *et al.* (2009) emphasises the requirement for pupils to formulate questions and provide answers during the inquiry process. Lai (2016), Lai *et al.* (2018) and Letina (2020) focused on inquiry based teaching.

The findings of these studies are mixed, and attitudinal outcomes explored are very diverse. Lin *et al.* (2009) indicated that the guided inquiry-based approach did not have a significant impact on pupils' perceptions of their classroom environment, including perceptions of teacher support, cooperation and engagement. In contrast, Polikoff *et al.* (2018) reported that the guided inquiry-based approach led to a significant increase in positive emotions such as excitement and a significant decrease in negative emotions such as feelings of boredom, frustration and confusion – but no effect on interest. Similarly, Chen *et al.* (2018) reported a significant positive effect on engagement in learning science for the experimental group compared to the control group, especially in terms of emotional engagement, enjoyment and pleasure. Letina (2020) found a positive effect on learning to learn competence of inquiry learning, defined as learning in which pupils plan research, make observations, descriptions, compare, collect, record, interpret and present data, draw conclusions, form assumptions, use literature and write reports. Lai (2016) and Lai *et al.* (2018) found positive effects of inquiry on attitudes, learning approach and self-regulation.

In terms of effects on specific groups of pupils, Lai *et al.* (2018) reported that pupils in the experimental group with initial higher self-regulation improve their learning achievement more than those with low self-regulation, with no differentiation in the control group.

Together these studies indicate that practical work and inquiry have a positive - or at worst, null - impact on pupils' affective engagement. The evidence on the effect on pupils' interest in science and their motivation to get involved in science is limited but seems to point at a lack of effect on these attitudinal aspects.

Critical thinking and argumentation

Six studies reported on outcomes related to critical thinking.

One small-scale study involving practical work reported an impact on critical thinking. Leuchter *et al.* (2014) reported greater conceptual progression for children using learning activities with practical work stations which focus on conceptual change compared with those taught the usual curriculum. This was evidenced by significant gains in quality of reasoning for the experimental group.

Five small-scale studies in the inquiry sub-cluster reported on the effect of an inquiry-based approach on pupils' scientific reasoning. The study by Chen and She (2014) exploring the effects of explicit reasoning prompts has been described above. In another study from Taiwan, Chen *et al.* (2019) used Modified Argumentation-Driven Inquiry (MADI) where the teacher specifically encourages an argumentation-based discussion leading to the explanations of the data collected. The control group experienced business-as-usual teaching including structured teacher-directed hands-on experimental work. Schalk *et al.* (2019) used a guided inquiry approach to design a set of teacher-directed structured practical tasks leading to 'conclusive results'.

The results in terms of the effects of inquiry on scientific reasoning of all four studies are positive. Schalk *et al.* (2019) reported a statistically significant positive effect on ability to apply the Control-of-Variables Strategy (CVS) using guided inquiry compared to direct teaching. Di Mauro and Furman (2016) indicated that experimental design skills significantly improved for an experimental group exposed to an inquiry curriculum but remained constant for the control group. Similarly, pupils using a Modified Argumentation-Driven Inquiry approach (Chen *et al.*, 2018) showed statistically significant improvements in argumentation skills compared to a control group using teacher-guided hands-on experiments in class. Positive effects were particularly prominent in the second semester, indicating that improvement in argumentation skills needs time. Taking an engineering design-based modelling approach, Li *et al.* (2016) found that pupils' (and in particular, boys) in the treatment group experienced significant gains in problem-solving ability when compared with pupils who did not analyse or optimise potential solutions.

In terms of what happens during inquiry in primary science, Chen and She (2015) found that pupils with explicit prompts outperformed the group who did not have to make their thinking explicit on scientific concept-dependent reasoning, and were able to generate a greater number of testable hypotheses, all at a level of statistical significance. Zhang (2018) investigated the effect of withholding answers during hands-on scientific investigations on pupils' scientific reasoning and found a positive effect on reasoning for pupils in the hands-on only condition (compared with 'hands-on and withholding answers' and 'direct instruction'), with the poorest performance when hands-on was combined with withholding information.

In conclusion, the findings of these studies indicate that inquiry can support primary school pupils' scientific reasoning and in particular the inquiry skills of experimental design and hypothesis formulation, although there is considerable diversity of approaches and specific outcomes measured.

Working scientifically and problem solving

Six small-scale quasi-experimental studies reported outcomes related to working scientifically, five of which investigated inquiry approaches, and one focused on practical work.

One study investigated the effect of practical work and reported an impact on working scientifically. Ünal and Aral (2014) found a large positive effect of a series of experiments which could be done by the children themselves on 'problem solving skills', although it is not clear what the experiment programme was compared to, as their 'routine educational process' was not described.

The five studies from the inquiry sub-cluster investigated different types of inquiry. Schalk *et al.* (2018) and DiMauro and Furman (2016) both investigated the impact of guided inquiry. DiMauro and Furman reported that the guided inquiry group, which experienced clear identification of dependent and independent variables, showed significant levels of improvement in experimental design performance at the end of the programme, with no differences for the control group. Similarly, Schalk *et al.* (2019) reported a statistically significant positive effect on ability to apply the Control-of-Variables Strategy (CVS) using guided inquiry compared to direct teaching, although it is important to note that the control group was not taught physics.

Another modified version of inquiry related to reasoning. Chen and She (2015) compared two inquiry programmes, one with an emphasis on reasoning (through prompts to encourage deductive and causal reasoning) and the other without. They found a small positive effect of inquiry which incorporated reasoning, noting that pupils in this group produced a significantly greater number of testable hypotheses, correct hypotheses, and correct evidence-based scientific explanations and a higher level of scientific reasoning than did the control group.

Lai *et al.* (2018) compared inquiry using the 5E instructional model with and without a self-regulation step, and found significantly higher information seeking tendency in the intervention (computer assisted self-regulation) group. Lin *et al.* (2009) were also interested in questions. They found that within the experimental group, the pupils who asked high-quality questions significantly outperformed pupils who asked low-quality questions in the design of experimental procedures but not in making hypotheses.

Taken together, these studies suggest a positive effect of various types of inquiry on a range of measures that indicate ability to work scientifically, although all were small studies, typically conducted in one or two schools. Even a more structured practical approach can support pupils to learn skills such as control of variables. There is limited evidence on the effectiveness of practical work as a conceptual change strategy.

Implementation and resources

None of the studies reported barriers to implementation, although access to resources and professional development were key for some of the strategies, and these are likely to require resourcing.

Costs were not reported, however there is a cost associated with resources used (such as Hot Wheels and LEGO) and half of the studies were funded externally.

Gaps in the evidence base

A number of approaches to practical work, inquiry and investigation were absent from the studies included in the review. These include Explorify and CREST (Creativity in Science and Technology) Awards (which have been evaluated with older pupils).

Another gap in the evidence base relates to outcomes. Whilst pupil attainment, and to a lesser extent attitudes, tended to be measured in studies in this cluster, more relevant outcomes might relate to thinking and working scientifically.

Whilst two studies on practical work included younger pupils, inquiry approaches were used exclusively with children aged eight or older, indicating a gap in understanding about the effects of inquiry on young pupils.

Quality of the evidence base

Using the quality assessment tool, one study was rated high-quality, with the others rated moderate (10) or low (7). A main influence on the quality assessment was study design (only two RCTs) and the lack of high-quality studies. Although there is consistency in findings, there is a small number of studies with varying approaches, with only two RCTs, and none conducted in England. Within studies, quality was affected by selection bias and the issue of confounders, in particular if there were important differences between groups prior to the intervention, was often not addressed specifically when describing the samples. All studies reported approaches that are generalisable to routine practice. There was some consistency in approach, but little in terms of outcome measures.

Applicability of the evidence base

Although the education systems are different in the countries where the studies in this cluster were conducted, all studies took place in classrooms using approaches that could be applied in England.

The inquiry approaches are applicable to the English education context, given their focus on scientific reasoning. Indeed, the 5E instructional model (on which several studies were based) is in use in schools in England. In the Swiss study (Schalk *et al.*, 2019), the authors noted two characteristics of their study which might limit applicability to England: the low number of welfare recipients (3%) and the autonomy teachers have to select what they teach. Similarly, the practical work cluster uses approaches already common in the English context, including Predict-Observe-Explain, design tasks and hands-on activities.

Overall evidence statement/Key findings

The evidence suggests that practical work and inquiry - particularly when guided or when there are prompts for reasoning – can have a positive effect, or a null effect, on primary school pupils' attainment in and attitudes towards science, their critical thinking skills and ability to work scientifically. However, there is limited evidence and further research is required in this area.

Table 22: Summary of the evidence on practical work, inquiry and investigation

Number of studies	18 (12 inquiry; 6 practical work)
Assessed quality of studies	High: 1 Moderate: 10 Low: 7
Location of studies	Argentina: 1 China: 1 Croatia: 2 Germany: 1 Serbia: 1 Switzerland: 2 Taiwan: 5 Turkey: 2 USA: 3
Outcomes of focus	Attainment (knowledge and conceptual understanding): 10 Attitudes, motivation and interest: 7 Critical thinking and argumentation: 5 Working scientifically and problem solving: 6
Design	Quasi-experiment: 16 Randomised control trial: 2 (inquiry and investigation)
Consistency of findings	Results were consistently positive when compared with business-as-usual, with the exception of two studies
Summary of the evidence base	• Study design: two randomised controlled trials • One high quality study • The majority of studies have very small samples • Variability in interventions and outcome measures. All studies except one used researcher developed instruments, rather than standardised measures • All interventions except two were developed and evaluated by the researchers reporting the work • All but two studies used different interventions and different instruments to assess outcomes

Part C: Conclusions and recommendations

Conclusions from the systematic review

In the introductory section of the review, it was noted that Slavin *et al.* (2014), who conducted a review of experimental evaluations of particular approaches in primary science, concluded that "The most important finding of the present review is the very limited number of rigorous experimental evaluations of elementary science programs." (p878). What has changed? The review by Slavin *et al.* focused on randomised control trials, and there are certainly more RCTs being undertaken now, together with an increased emphasis on experimental designs in evaluations. Well-formulated RCTs and experimental studies contribute to the knowledge base on 'what works?'. However, this review highlights that several challenges remain in building that evidence pipeline.

The systematic review has generated a very broad-ranging and diverse range of evidence. It demonstrates that there is wide-spread interest across a range of countries in the effective teaching of science in primary schools.

Indicators of the nature of the interest come from both topics on which the studies focus, and from looking at where Governments and other external funders (such as charitable groups) are supporting research. There is clearly considerable interest in inquiry approaches, and the role of language learning, pupil writing, and pupil talk in science lessons. Areas in which Governments and other groups are funding work include pupil talk (as part of dialogic teaching), group work, and the use of software to support science teaching. Surprisingly, given the importance of assessment, there were very few studies in this area that met the criteria for inclusion in the review.

In relation to the focus of the studies, many explored more than one of the areas in the clusters identified in the review. For example, studies that focused on language development often had links to writing skills, those that focused on explicit instruction did so in a context of teaching through inquiry, and critical thinking approaches often had connections with co-operative and collaborative approaches and learning through talk.

It is very rare for researchers not to be the evaluators of their interventions. Exceptions include the studies by Alexander (2018); Hanley *et al.* (2015; 2018) and Kitmitto *et al.* (2018), on dialogic teaching and Thinking, Doing, Talking Science. It is also very rare for null or negative results to be reported. One exception is the study by Hand *et al*. (2018) on using their Science Writing Heuristic to develop argumentation skills. Few studies used an approach in a range of scientific topic contexts (see Alexander, 2018 for an example of a study which does).

Studies are less likely to use externally developed and validated instruments in the evaluation of interventions. By far the most common method is for researchers to use instruments they have developed themselves. This has two potential consequences. Firstly, it may compromise the validity of the research in the individual studies, leading to enhanced effect sizes. Secondly, for research in the same area, it militates against the building up of a cumulative evidence base and the conducting of meta-analyses of findings.

There are numerous examples of individual studies which meet criteria of rigour and high-quality in themselves but are focused on very limited or niche areas, such as the impact of specific ICT applications; they use measures very close to the intervention, and use weak or ill-defined controls.

What are the most effective approaches for improving outcomes in primary science?

Given the inconsistency of focus, outcomes and measures used, it is not meaningful to rank or compare approaches within or between clusters.

Language and literacy-based approaches in primary science were associated with positive effects across a range of outcomes including attainment and scientific language. Approaches included dialogic teaching (Alexander, 2018, Jay *et al.*, 2017), science and language integrated instruction (Cervetti *et al*., 2012; Connor *et al.*, 2017; Goldschmidt & Jung, 2011; Kim *et al*., 2021; Vitale & Romance (2012) and science for language learners (Llosa et al., 2016; Maerten-Rivera *et al.*, 2016). Integration of science and literacy over an extended period can improve cognitive and affective outcomes for pupils in science and literacy, particularly for pupils who are less proficient in the language. These studies typically took place over an extended period and involved substantial professional development for teachers. Indeed, changing the nature and amount of CPD is one of the factors that may account for the inconsistency in findings for Thinking, Doing, Talking Science (Hanley *et al.*, 2015; Hanley *et al.*, 2020; Kitmitto *et al.*, 2021). The link between language, thinking and science is common in a number of approaches (Cervetti *et al.*, 2012; Connor *et al*., 2017; Hanley *et al.*, 2015; 2020; Llosa *et al*., 2016; Maerten-Rivera *et al*., 2016). Together, these studies suggest that increased talk enables pupils to identify prior knowledge and evidence to support their thinking, integrate new ideas, practise articulating these and receive feedback from peers and teachers. Connections exist to high quality studies in other clusters. For example, in the explicit instruction and related approaches and practical work and inquiry clusters, explicit teaching of scientific vocabulary through guided inquiry (Doabler *et al*., 2021; Schalk *et al.*, 2019) produced some positive effects, as did the use of the 5E instructional strategy involving the phases of engage, explore, explain, elaborate, evaluate (Polikoff *et al.,* 2018).

Although the assessment cluster consisted of a small number of studies, findings were consistent, with positive effects on attainment (science competence) and attitudes (motivation and perceived competence) reported. Approaches explored included embedding formative assessment in teaching (Decristan *et al.*, 2015a;b; Hondrich *et al.*, 2018), guided peer feedback (Hwang *et al.*, 2018) and bidirectional peer assessment (Hwang & Chang, 2021), all of which create opportunities for meaningful classroom interactions focused on the learning goal, and time for thinking and talking, consistent with the findings from the language, literacy and text-based approaches cluster above.

A range of approaches in the cooperative and collaborative approaches cluster reported positive effects on attainment and attitudes, including collaborative writing to older peers (Chen *et al.*, 2013), discussion (Chang & Hsin, 2021), online collaboration (Looi *et al.*, 2010) and collaboration during differentiation (Eysink *et al.*, 2017). Collaboration was also a feature of studies in the ICT cluster (Hwang *et al.*, 2020), and language and literacy cluster (Alexander, 2018, Jay *et al.*, 2017). Inconsistency in this cluster was a result of two studies using the science writing heuristic (Hand *et al.*, 2018; Reeves *et al.*, 2013). Both studies found null effects on attainment, but a positive effect on critical thinking was calculated for the use of the science writing heuristic (Hand *et al.*, 2018).

Approaches to learning outside the classroom included short- and long-duration from field visits (Glick & Samarapung, 2008),field work (Scott & Boyd, 2016), visits to science centres and interactions with scientists (Mills & Katzman, 2015), and use of the local habitat and school grounds (Wells *et al.*, 2015). Consistently positive effects were reported on attainment, with small positive or null effects on a range of attitudinal outcomes. These approaches allow for collaboration, experiential learning and time outdoors and engagement with nature. There are also connections with the study by Burt *et al.* (2022) in the context-based and cross-curricular cluster, which reported on the use of a climate change curriculum incorporating aspects of learning outside the classroom in the use of hydroponic gardening, worm composting and use of rainwater catchment systems. Positive effects of this approach on attainment in state-wide tests were calculated.

Finally, there was evidence of some positive effects across context-based and cross curricular approaches, critical thinking and argumentation, explicit instruction and related approaches, ICT supported and online approaches and practical work and inquiry approaches.

What moderating factors, barriers and facilitators influence the effectiveness of approaches?

This section considers the moderating factors that influenced the effectiveness of the approaches. A number of patterns emerged from the review.

A frequent moderating factor was the provision of CPD for teachers, and whole school support for the approach and associated CPD. For a number of the high-quality studies (e.g. Alexander, 2018, Hand *et al*., 2018; Hanley *et al.*, 2015; 2018; Maerten-Rivera *et al.*, 2016), it was clear that an extensive CPD programme was needed to support the intervention. It may be that one option in any 'proof-of-principle' (efficacy) trials would be to compare an in-depth version of any CPD with a 'lighter-touch' version to get an indication of the amount of CPD required.

A second strong moderating factor was the length of the intervention. Again, a characteristic of the high-quality interventions was that they were of several months or one-to-two years' duration.

In terms of the nature of the interventions, approaches which had a positive impact on pupil outcomes often included opportunities for pupils to share their thinking with peers or teachers.

Taking a step back from the studies themselves, it seems very likely that the researchers were themselves moderating factors in many of the studies. They developed the interventions and, in almost all cases, undertook the evaluation of the interventions. Moreover, the instruments used to assess the effectiveness of the interventions were most commonly developed by the researchers themselves.

What are the differential impacts of the approaches on identified disadvantaged pupils?

Few studies reported differential outcomes for different groups of pupils. Only the statistically significant impacts are reported here, with conclusions drawn from all study clusters.

Pupils eligible for free or reduced school meals

Approaches which show greatest promise for improvement of outcomes of pupils eligible for free or reduced school meals include approaches which provide opportunities to make thinking explicit through words, whether spoken or written.

There were more pronounced positive effects for pupils eligible for free or reduced meals when pupils were asked to use writing for argumentation. For example, using a science writing heuristic, an approach to guided inquiry which scaffolds pupils' argumentation about laboratory investigations (Hand *et al.*, 2018) or writing letters to construct an argument using questions, claims, data and evidence (Chen *et al.*, 2013).

Similarly for talk, pupils eligible for free school meals made additional progress during an extended dialogic teaching intervention (Alexander, 2018). This finding was supported by the external evaluation report by Jay *et al.* (2017) who found that in schools using Alexander's dialogic teaching approach, pupils eligible for free school meals made two additional months' progress in science compared to FSM children in control schools. However, they noted that the smaller number of pupils eligible for free school meals in the trial limits the security of this result. There were indications of particularly positive effects of Thinking Doing Talking science in earlier studies (Hanley *et al.*, 2015; 2020) but not in a later study with a different model of CPD for teachers in place (Kitmitto *et al.*, 2018).

Pupils with low prior attainment

Several studies reported differential outcomes for children with low prior attainment. These were not always positive, for example, an extended (one term) duration intervention which involved a Mars-colonisation themed STEAM intervention reported non-significant gains for low achievers compared to significantly increased learning outcomes for high achievers.

Strategies described in studies reporting additional benefits for children with low prior attainment included the use of the 'self-explain, discuss, re-explain' strategy (Chang & Hsin, 2021), which improved outcomes particularly for low achievers (but with a ceiling effect for high achievers) and Thinking, Doing, Talking Science, which had a particularly positive impact on children with low prior attainment, particularly girls (Hanley *et al.*, 2015). In common with those reporting gains for children eligible for free or reduced meals, these approaches emphasise making scientific thinking explicit through words, whether spoken or written.

Special educational needs

Several studies reported gains for children with special educational needs.

Guided inquiry shows promise for children with additional learning needs. The science writing heuristic reported by Reeves *et al.* (2013) reported increasingly beneficial effects with increasing exposure for children with special educational needs. Similarly, the 'Science Explorers' curriculum (Doabler *et al.*, 2021) unit which involves activation of prior knowledge, explicit instruction of vocabulary, read-alouds and investigation activities showed potential for improving outcomes for children with or at risk of developing difficulties with learning.

Secondly, in the assessment cluster, an adaptation involving audible image description was found to enhance attainment for visually impaired children who use Braille, but not for those who use print (Ferrell *et al.*, 2017).

English Language Learners

Approaches which reported specific benefits for pupils who speak English as an additional language include scaffolded instructional discourse (Decristan *et al.*, 2015a) and instruction which integrates scientific content with literacy instruction (Connor *et al.*, 2017; Llosa *et al.*, 2016; Maerten-Rivera *et al.*, 2016; Zwiep & Straits, 2013). There was some evidence in the qualitative data presented by Kim and Olaciregui (2008) which suggest that how materials are presented can support conceptual understanding, with concept maps made using digital platforms being more effective than more traditional text-based layouts, likely due to the way in which connections can be represented and manipulated.

What does the review contribute to high-quality science teaching in primary schools?

The review presents a synthesis of evidence that can provide information on what has been measured to be effective, but not necessarily what is important. There are important gaps in the evidence base. For example, a number of commonly used and/or research-informed approaches and have not yet been evaluated using research designs using a counterfactual. These may or may not be effective across a range of outcomes. Furthermore, issues with the quality of primary science teaching need to be resolved by considering the impact of interventions in the context of the broader challenges facing primary science teaching, and it is clear from what has been written (e.g. most recently, Ofsted Research Review of Science, 2021; Wellcome, 2020; Bianchi *et al.,* 2021) that these challenges are considerable.

Provision is very variable, meaning that a key first step in improving primary science teaching is to create a basic expectation for science provision and ensure that this is being met. In this context, it is important to note that one of the main aspirations for primary science associated with the removal of testing at 11 has not been achieved. Rather than, as hoped, giving teachers more flexibility to be creative in their teaching of science in preference to 'teaching to the test', less science is now being taught.

As Ofsted and others have noted, too much emphasis is currently placed on individual activities that pupils might enjoy, and not enough on planning a coherent curriculum. Thus, a second key step is to decide what sort of curriculum should be followed in primary schools. The review points to positive impacts on pupil outcomes from curricula which integrate inquiry, language and other contexts.

Once there is clarity over the status and nature of the curriculum, the findings of the systematic review need to be considered, and particularly the findings of high-quality interventions identified as having a positive effect. Any intervention needs to be located within the framework of the curriculum in order to reduce the chances of individual activities in an intervention being selected on the basis that pupils will find them 'enjoyable'.

Evidence from the review (particularly from studies judged to be of high-quality) points to successful interventions requiring considerable time, and a substantial component of CPD being required for interventions. The CPD is required prior to implementing the intervention, and during the intervention itself. This carries with it the clear message that teachers teaching science in primary schools need considerable support over an extended period of time to develop the knowledge, confidence and skills to teach high-quality science. A recent report found that science provision on programmes of initial teacher education can be as little as 7.5 hours over a year (Wellcome, 2017), which contrasts with the extensive CPD supporting some of the interventions report. This suggests that there are unlikely to be many, if any, 'quick fixes'.

The review also indicates that effective change in the classroom is more likely to occur if teachers are provided with structured resources to use with their pupils, and the support to use the resources in lessons during the implementation phase. Such resources need to be developed and located within a whole-curriculum framework.

The review identifies some approaches and/or interventions that seem likely to bring about improvements, including those that appear to be worth subjecting to trials in the English context to establish proof of principle.

The findings of this review resonate with the (few) recent reviews of primary science.

In common with Potvin & Hasni (2014) whose review focused on motivation, interest and attitudes, we found a body of literature consisting of a high number of studies reporting positive results, often using bespoke measurement instruments to assess the impact of interventions, and with poor articulation of what happened in the control group. Their conclusion - that effort usually produces effect when it comes to interest, motivations and attitudes – is broadly supported by the studies in this review.

We found some evidence of positive impact of integrated curricula in common with Gresnigt *et al.* (2014). Following an assessment of evidence on the effectiveness of different approaches to integrating science across the primary curriculum, Grenigt *et al.* (2014) proposed a model suggesting that the more transdisciplinary a curriculum was, the greater the need for teacher commitment, professional development, teacher support and sustained facilities – but that this investment supports the development of more positive attitudes, greater teacher commitment and better development of higher order thinking skills. Similarly, our review found that CPD was often an essential ingredient of successful curriculum interventions.

In common with a review of primary science programmes (Slavin *et al.*, 2014), we found positive impacts of inquiry programmes, particularly the 5E model used in a STEM context (Polikoff *et al.*, 2018) and Science Explorers guided inquiry (Doabler *et al.*, 2021).

Finally, this review is consistent with a couple of reviews which point to the effectiveness of language and literacy-based approaches to improve science outcomes. Huerta and Garza (2019) found support for writing-based strategies and Nunes *et al.* (2017) found that approaches to improving literacy are generally effective in supporting the science learning of pupils from low socio-economic status backgrounds. The present review adds to the evidence, finding that approaches to literacy teaching (such as scaffolded instructional discourse, content and language integrated instruction, and explicit instruction of vocabulary) and the deliberate use of pupil talk (for example, through dialogic teaching approaches or Thinking, Doing, Talking Science) can benefit all pupils, particularly those who speak English as an additional language.

Applicability of theoretical and practice review

The applicability of the review findings is now discussed in relation to teachers' approaches to the curriculum, teaching and assessment as identified in the review of practice.

Teachers in the focus groups described science being planned into the curriculum in four main ways: as part of thematic units or topics; driven by scientific knowledge; driven by 'working scientifically' and integrating knowledge and working scientifically. Whilst no studies in the review compared these four approaches, there was some evidence to suggest a positive impact of cross-curricular approaches including a cross-curricular climate change curriculum (Burt *et al.*, 2022), school garden curriculum (Wells *et al.*, 2015) and a STEM speedometry curriculum (Polikoff *et al.*, 2018) on attainment and attitudes towards science. A larger body of evidence indicates that curricula which integrate science, inquiry and literacy have a positive impact on science conceptual understanding and English attainment, suggesting that these integrated approaches are beneficial to all of the subjects included. Primary schools (or groups of schools in an academy) in England tend to develop their own curriculum rather than rely on published schemes, meaning that that practice is likely to vary a lot from school to school, making a more challenging context for experimental studies.

The review of practice revealed an appetite for evidence on how to assess primary science, with a variety of approaches used in schools, including self-assessment, peer-assessment and teacher assessment. The review found relatively few studies on assessment using an experimental design. That said, the evidence suggests better attainment outcomes for pupils in intervention groups using formative assessment compared to control groups. A range of resources were identified as valuable by teachers, including the Teacher Assessment in Primary Science (TAPS) resources developed by staff at Bath Spa University, which are currently being evaluated in a randomised control trial (Focus4TAPS) funded by the Education Endowment Foundation.

The review of practice found teachers referring to a number of approaches to teaching science. Some of these are well supported by the evidence base, including approaches to teaching language explicitly in science. Other approaches were absent from the evidence base, including those based on pupils' preconceptions. Practical work (rather than inquiry) was a widely referenced approach, and an area where the review found relatively few studies, and only of moderate-quality to support its use to bring about gains in conceptual understanding. It is important to recognise that teachers may select approaches with different, or multiple aims in mind.

Finally, teachers in the review of practice described the importance of involving pupils in whole school level science opportunities including as science ambassadors and science fairs, working with visitors including members of the local community and STEM ambassadors, and approaches linked to outdoor learning including gardening, outdoor investigations and participating in forest schools. The review found evidence to support opportunities to learn outside the classroom and outdoor investigations, but none specifically focusing on forest schools.

The biggest barrier to implementation in English classrooms is the time required to carry out the interventions in a meaningful way. Curriculum time for science has reduced since Key Stage 2 tests for science were abolished (Ofsted, 2021). In the review of practice, constraints that teachers identified as influencing their practice included the national curriculum, Ofsted, school funding and resourcing, school leadership, assessment policies and access to professional development, so the stated position of these bodies in relation to teaching approaches is likely to act as a facilitator or barrier.

A common approach amongst interventions in the review was to draw on psychological theories, particularly in relation to learning and memory, but also in relation to motivation and interest. Sociological theories (e.g. science capital, critical theory and capabilities approach) were uncommon as a basis of interventions. Studies often synthesised theory and existing evidence to describe bespoke logic models or theories of change for the intervention.

Implications and recommendations

Implications for policy and practice

This review set out to answer the question, 'What approaches are most effective to improve pupil outcomes in primary science, in what context, and how?' The evidence base that has emerged from the review is characterised by variability in the quality of the studies undertaken in a range of key areas. Moreover, a number of interventions are untested through causal research designs in England. There is also a lack of clarity in a number of studies about the precise nature of the study design in that interventions are generally explained in some detail, but little is said about how the experience of the control groups and intervention groups differ. Thus, at this point, the evidence base itself does not lend itself to many firm recommendations for policy or practice on the value of specific approaches. Rather it points to the need for an agenda that would seek to put into place a high-quality research evidence base, aligned to the purpose of primary science education. A key aspect of the agenda involves making decisions about the aims and structure of the primary science curriculum in terms of which concepts and contexts to prioritise, and in which order they should be taught to ensure that there is a logical progression in pupils' understanding. It is also important to consider what primary-aged children should learn about the nature and culture of science and how to practise science. This should take account of the work done on 'big ideas' in science, together with relevant research on conceptual development in science and research on the ideas in science known to cause pupils difficulties.

With a clear purpose and curriculum structure in place, it is then possible to look at teaching approaches drawing on the evidence from interventions such as those explored in this systematic review. There is good evidence to suggest that talking and writing about science provides opportunities for pupils to think about and verbalise their ideas which, in turn, helps develop their understanding of science. The evidence base from studies undertaken in England is small, but points to the effectiveness of interventions seeking to increase the amount of pupil dialogue in lessons. There are also studies in other countries which appear to offer valuable insights into ways of developing pupils' writing skills. Other approaches which might be considered include formative assessment, cooperative and collaborative approaches and opportunities to learn outside the classroom, even if only using school grounds.

Considerable attention is currently directed at inquiry-based learning and explicit (direct) instruction, following observations made in the PISA (Programme or International Student Assessment) findings in 2016 (OECD, 2016) which prompted debate about the relative merits of direct instruction and inquiry-based approaches. The evidence base on explicit instruction is small but points to benefits in instruction on scientific vocabulary and teaching experimental design. There is a more solid evidence base indicating that guided inquiry can have a positive impact on pupils' attainment in science and ability to work scientifically. Interestingly, inquiry and explicit instruction were often used together in studies included in the review. It would seem sensible to identify a small number of interventions that could be subjected to 'proof-of-principle' studies, particularly where the evidence on the approach comes from studies outside England. This would, in turn, provide an evidence base for making decisions about approaches to recommend in the teaching of primary science.

The review also provides evidence of two key matters that extend beyond individual approaches. Firstly, teachers need support to implement interventions, and time is needed for changes in practice to take place. Secondly, there needs to be more consistency over measures used to assess the effectiveness of interventions, including making more use of existing validated instruments and of external evaluation of interventions.

It is apparent in the review that the majority of studies focused on attainment. It is important to be aware of the distinction between (i) outcomes that can be (and are) measured, (ii) outcomes that are desirable to achieve, but might be more difficult to measure, and (iii) those outcomes which are probably unmeasurable, e.g. those which value human relations (and human relations with the planet). There is a need to prioritise outcomes of importance beyond attainment, using well-established measures. Whilst many of the studies included in the review focus on rather narrow measures of attainment, teachers we spoke to in focus group discussions had a much broader ambition for school science and the role that it can play in developing positive dispositions towards learning, building curiosity, thinking better, and securing more equitable representation in science in the future.

The review certainly points to the need for 'more research'. Whilst research is often criticised for concluding that more research is needed, the review identifies areas on which to focus. A call for 'more research' places demands on funders, researchers, teachers and young people. Participation in research studies must be adequately resourced for all stakeholders. Additionally, it will be important to listen to stakeholders (teachers, the science education community) when making decisions about which approaches to prioritise, and setting a research agenda for primary science.

Decisions on these areas could be usefully informed by the wide range of outcomes identified as important by teachers in the review of practice.

Teachers and science subject leaders may wish to reflect on their use of the approaches which the review suggests may be promising for improving a range of outcomes of primary science. These include approaches at the level of the curriculum, for example, cross-curricular and context-based units and the integration of science and literacy instruction, particularly for children who speak English as an additional language. There is good evidence to suggest that the integration of science and literacy improves outcomes for both subjects.

Pedagogical approaches that may be promising for improving outcomes include formative assessment, the use of inquiry supported by prompts and explicit instruction of scientific vocabulary and inquiry skills, experiences of learning outside the classroom and the use of ICT to support learning outside the classroom, inquiry and collaboration.

Implications for research

This systematic review points to three ways in which the evidence-base might be considered to have gaps in it.

The first relates to the gap between what approaches have been tested, and what approaches are used by teachers in England. For example, the review of practice indicated four approaches to the primary science curriculum being used by teachers: a thematic approach, knowledge-based approach, an approach driven by working scientifically and an integrated approach. We found no studies with a causal design comparing these curricular approaches. Nor did we find studies using teaching approaches identified in the review of practice, for example identification of pupils' preconceptions of scientific concepts, participation in science fairs and being a scientific ambassador, or the science capital approach. It is therefore important to take account of these gaps in the evidence base when planning a research agenda and possible trials of interventions in the future.

The second gap in the evidence base concerns the lack of cohesion in the evidence-base. Well-constructed experimental studies are necessary but not sufficient to provide a sound evidence base. They need to be located within a framework of which areas are deemed to be particularly worthy of study.

The third gap in the evidence base results from the lack of consistency over how interventions should be evaluated, with the preponderance of studies being evaluated by the researchers developing the intervention, and with researcher-developed instruments, often in a single topic area. Whilst there may well be good reasons for using instruments developed for a particular intervention, this should be accompanied by the use of more standardised, externally developed instruments to gather information on student attainment, attitudes and other outcomes of interest for primary science. It may be that some of the items from PISA and TIMSS surveys provide a way ahead here, and a standardised attainment test for primary science has been developed by the EEF (EEF, 2022). There may also be a need to establish a bank or repository of such instruments, as has been the case in Modern Foreign Languages (MFL) education (the IRIS digital repository: https://www.iris-database.org/)

Both the second and third of the gaps described above also contribute to lack of depth in the evidence base. There is little sense of cumulative knowledge being built up in key areas such as practical work, which is a particular concern given its importance and prevalence in practice (Ofsted, 2021; Wellcome Trust, 2019; Appendix 2).

At a more detailed level, there is an issue with a number of studies where there is a lack of information about the experiences of control groups in experimental studies. Sometimes it is clear that it is 'business as usual' (i.e. an intervention is compared with the standard curriculum pupils would have experienced), whereas on other occasions this does not appear to be the case. Few studies seem to have 'strong' controls where only the approach is being tested. Greater clarity in reporting is needed in these cases. The tools for assessing quality of studies predetermined in this review come from a tradition of research in medicine. The limited applicability of these tools to educational contexts points to the need for the development of a tool for assessing quality of educational studies.

The review of practice suggests that evidence would be useful in a number of approaches to curriculum, teaching and assessment, addressing the following questions:

- Curriculum: what is the most effective way to integrate working scientifically and science content into the school curriculum?
- Teaching: what are the most effective ways of: (i) teaching using practical work, (ii) bringing about conceptual change, (iii) participating in science fairs and engaging with science ambassador programmes, (iv) promoting learning outside the classroom, particularly forest schools, (v) leading science in schools, and (vi) improving pupil outcomes through teacher CPD?
- Assessment: which approaches to summative assessment have a positive impact on pupils' outcomes (including non-cognitive outcomes)?

Limitations

The review included only studies published and containing data collected between 2007 and 2021 and published in English. There may be high-quality studies published prior to 2007 and/or in languages other than English which assess the effectiveness of approaches to teaching primary science. Whilst the review is current, bibliometric studies (Chang, Chang & Tseng, 2010; Li, Wu & Tsai, 2009) suggest that research on conceptual change and concept mapping declined in the 2000s, along with declining attention to cultural, social and gender issues, indicating the possibility of absences based on themes that have been less popular in recent decades.

Quality was assessed following screening of studies. We did not use quality of study as an inclusion criterion. This means that the review was inclusive of studies with a causal research design, but many of these were small in scale and the highest-quality standards were not always met. For example, some studies omitted descriptions of the nature of the comparison group and teaching approach for the control group.

Studies were allocated to clusters for synthesis, and steps were taken to ensure that this was meaningful and that appropriate approaches were grouped together. However, the clusters reported represent just one way of organising studies. As it would be appropriate to consider some reports in a number of clusters, we have drawn our conclusions at a more fine-grained level.

The review did not include studies that focused only on the early years foundation stage, although children of this age were included in some studies reviewed. This is a crucial stage for children because of the rapid developmental changes that occur during these years. However, the very different context, aims and focus of education during the early years (science is not mentioned in the early years foundation stage framework for England) mean that the review will have limited applicability to teachers of the Reception year in primary schools in England.

Finally, the predominance of statistically significant positive results of approaches used in interventions suggests some publication bias.

Proposed recommendations for the Guidance Panel

The review team has followed the systematic review procedure as carefully as possible to try to ensure that its conclusions on the evidence base are transparent and valid. The team is conscious that some element of professional judgement is inevitable is presenting the findings of any research, including these from a systematic review. This is particularly the case when looking to identify the conclusions which the evidence from the review does and, most crucially, does not allow one to conclude. The review team is of the view that the evidence base on the effectiveness of approaches to the teaching of primary science is limited, though with some areas that demonstrate considerable potential (language and literacy approaches; co-operative and collaborative approaches; formative assessment; learning outside the classroom). The review team is also of the view that the findings of the review cannot be taken as the single starting point for making recommendations for practice, or setting a broader research agenda for improving primary science teaching.

We recommend that the guidance panel is clear about their shared understanding of the purpose of primary science, and what outcomes are desired for pupils. This shared understanding should be informed by the research on practice (Appendix 2) and used to identify priority approaches and research studies which are likely to be useful to teachers in achieving those aims.

Team

The team is listed below in alphabetical order, along with strands each member was involved in.

- Dr Lucy Atkinson, Research Fellow, University of York. Systematic review.
- Professor Judith Bennett, Co-Principal Investigator, University of York. Study design, systematic review.
- Sarah Compton, Project Administrator, University of York. Review of practice, evidence map, systematic review.
- Dr Lynda Dunlop, Co-Principal Investigator, University of York. Study design, review of practice, evidence map, systematic review.
- Dr Helen Glasspoole-Bird, Research Fellow, University of York. Review of practice, evidence map, systematic review.
- Dr Pam Hanley, Research Fellow, University of York. Study design.
- Dr Fred Lubben, Research Fellow, University of York. Evidence map, systematic review.
- Professor Michael J. Reiss, University College London Institute of Education. Theoretical review.
- Dr Maria Turkenburg-van Diepen, Research Fellow, University of York. Review of practice, evidence map, systematic review.

Timeline

Date	Activity
June 2021	Review begins Ethical approval for research on practice submitted
July 2021	Ethical approval granted Guidance panel meeting
August 2021	Protocol (first draft) submitted to EEF
October 2021	Research on practice: focus groups held
November 2021	Research on practice submitted to EEF Protocol (final draft) and addendum 1 published
December 2021	Outcomes of interest for systematic review determined Search strings developed following review of practice
January 2022	Search strings agreed Searching begins
February 2022	Searching ends Screening begins Format of evidence map determined
March 2022	Light touch data extraction tool approved
April 2022	Evidence maps produced
May 2022	Evidence map presented to EEF and guidance panel Data extraction tool approved Data extraction, synthesis and quality appraisal begins Theoretical review completed
June 2022	Addendum 2 to the protocol published Data extraction, synthesis and quality appraisal
July 2022	Decision on meta-analysis made: significant heterogeneity means that there will be no meta-analysis
August 2022	Report of systematic review drafted
September 2022	Systematic review submitted to EEF
October 2022	Revised report of systematic review submitted to EEF
November 2022	Guidance panel meeting
Ongoing	Support to produce guidance report

Conflicts of interest

The review team has no conflicts of interest to report. Hanley is author of several studies included in the review but was not involved in screening or reviewing.

Registration

The review protocol and addenda are published at: https://educationendowmentfoundation.org.uk/education-evidence/evidence-reviews/primary-science.

References of studies included in the systematic review

Alexander, R. (2018). Developing dialogic teaching: genesis, process, trial. *Research Papers in Education*, *33*(5), 561–598. https://doi.org/10.1080/02671522.2018.1481140

Arias, A. M., Smith, P. S., Davis, E. A., Marino, J.-C., & Palincsar, A. S. (2017). Justifying Predictions: Connecting Use of Educative Curriculum Materials to Students' Engagement in Science Argumentation. *Journal of Science Teacher Education*, *28*(1), 11–35. https://doi.org/10.1080/1046560X.2016.1277597

Barak, M., & Dori, Y. J. (2011). Science Education in Primary Schools: Is an Animation Worth a Thousand Pictures? *Journal of Science Education and Technology*, *20*(5), 608–620. https://www.jstor.org/stable/41499426

Baumfalk, B., Bhattacharya, D., Vo, T., Forbes, C., Zangori, L., & Schwarz, C. (2019). Impact of model-based science curriculum and instruction on elementary students' explanations for the hydrosphere. *Journal of Research in Science Teaching*, *56*(5), 570–597. https://doi.org/10.1002/tea.21514

Berry, J., Potter, J., & Hollas, V. (2013). Concept maps and informational read-alouds: strengthening both science and reading for elementary students. *Texas Journal of Literacy Education*, *1*(1), 29–44.

Bigozzi, L., Vezzani, C., Tarchi, C., & Fiorentini, C. (2011). The role of individual writing in fostering scientific conceptualization. *European Journal of Psychology of Education*, *26*(1), 45–59. https://www.jstor.org/stable/23883635

Bravo, M. A., & Cervetti, G. N. (2014). Attending to the Language and Literacy Needs of English Learners in Science. *Equity & Excellence in Education: University of Massachusetts School of Education Journal*, *47*(2), 230–245. https://doi.org/10.1080/10665684.2014.900418

Burt, K. G., Burgermaster, M., D'Alessandro, D., Paul, R., & Stopler, M. (2020). New York City fourth graders who receive a climate change curriculum with hydroponic gardening have higher science achievement scores. *Applied Environmental Education and Communication*, *19*(4), 402–414. https://doi.org/10.1080/1533015X.2019.1611504

Cervetti, G. N., Barber, J., Dorph, R., Pearson, P. D., & Goldschmidt, P. G. (2012). The impact of an integrated approach to science and literacy in elementary school classrooms. *Journal of Research in Science Teaching*, *49*(5), 631–658. https://doi.org/10.1002/tea.21015

Chang, T.-L., & Hsin, H.-T. (2021). The effect of the Self-explain–Discuss–Re-explain (SDR) learning strategy on high- and low-achieving fifth-grade students' achievement in science. *Research in Science & Technological Education*, *39*(4), 461–488. https://doi.org/10.1080/02635143.2020.1772223

Chen, C.-T., & She, H.-C. (2015). The effectiveness of scientific inquiry with/without integration of scientific reasoning. *International Journal of Science and Mathematics Education*, *13*(1), 1–20. https://link.springer.com/article/10.1007/s10763-013-9508-7

Chen, H.-T., Wang, H.-H., Lu, Y.-Y., Lin, H.-S., & Hong, Z.-R. (2016). Using a modified argument-driven inquiry to promote elementary school students' engagement in learning science and argumentation. *International Journal of Science Education*, *38*(2), 170–191. https://doi.org/10.1080/09500693.2015.1134849

Chen, H.-T., Wang, H.-H., Lu, Y.-Y., & Hong, Z.-R. (2019). Bridging the Gender Gap of Children's Engagement in Learning Science and Argumentation Through a Modified Argument-Driven Inquiry. *International Journal of Science and Mathematics Education*, *17*(4), 635–655. https://link.springer.com/article/10.1007/s10763-018-9896-9

Chen, Y.-C., Hand, B., & McDowell, L. (2013). The effects of writing-to-learn activities on elementary students' conceptual understanding: Learning about force and motion through writing to older peers. *Science Education*, *97*(5), 745–771. https://doi.org/10.1002/sce.21067

Cheng, M.-C., Chou, P.-I., Wang, Y.-T., & Lin, C.-H. (2015). Learning effects of a science textbook designed with adapted cognitive process principles on grade 5 students. *International Journal of Science and Mathematics Education*, *13*(3), 467–488. https://link.springer.com/article/10.1007/s10763-013-9471-3

Cohen, M. T., & Johnson, H. L. (2012). Improving the acquisition and retention of science material by fifth grade students through the use of imagery interventions. *Instructional Science*, *40*(6), 925–955. https://link.springer.com/article/10.1007/s11251-011-9197-y

Connor, C. M., Dombek, J., Crowe, E. C., Spencer, M., Tighe, E. L., Coffinger, S., Zargar, E., Wood, T., & Petscher, Y. (2017). Acquiring Science and Social Studies Knowledge in Kindergarten Through Fourth Grade: Conceptualization, Design, Implementation, and Efficacy Testing of Content-Area Literacy Instruction (CALI). *Journal of Educational Psychology*, *109*(3), 301–320. https://doi.org/10.1037/edu0000128

Cvjeticanin, S., Obadovic, D., & Rancic, I. (2015). The Efficiency of Student-led and Demonstration Experiments in Initial Physics-Chemistry Education in Primary School. *Croatian Journal of Education - Hrvatski časopis Za Odgoj I Obrazovanje*, *17*, 11–39. https://hrcak.srce.hr/en/file/218235

Dankenbring, C., & Capobianco, B. M. (2016). Examining Elementary School Students' Mental Models of Sun-Earth Relationships as a Result of Engaging in Engineering Design. *International Journal of Science and Mathematics Education*, *14*(5), 825–845. https://link.springer.com/article/10.1007/s10763-015-9626-5

Decristan, J., Hondrich, A. L., Büttner, G., Hertel, S., Klieme, E., Kunter, M., Lühken, A., Adl-Amini, K., Djakovic, S.-K., Mannel, S., Naumann, A., & Hardy, I. (2015b). Impact of Additional Guidance in Science Education on Primary Students' Conceptual Understanding. *The Journal of Educational Research*, *108*(5), 358–370. https://doi.org/10.1080/00220671.2014.899957

Decristan, J., Klieme, E., Kunter, M., Hochweber, J., Büttner, G., Fauth, B., Hondrich, A. L., Rieser, S., Hertel, S., & Hardy, I. (2015a). Embedded Formative Assessment and Classroom Process Quality: How Do They Interact in Promoting Science Understanding? *American Educational Research Journal*, *52*(6), 1133–1159. https://doi.org/10.3102/0002831215596412

Di Mauro, M. F., & Furman, M. (2016). Impact of an inquiry unit on grade 4 students' science learning. *International Journal of Science Education*, *38*(14), 2239–2258. https://doi.org/10.1080/09500693.2016.1234085

Doabler, C. T., Therrien, W. J., Longhi, M. A., Roberts, G., Hess, K. E., Maddox, S. A., Uy, J., Lovette, G. E., Fall, A.-M., Kimmel, G. L., Benson, S., VanUitert, V. J., Emily Wilson, S., Powell, S. R., Sampson, V., & Toprac, P. (2021). Efficacy of a Second-Grade Science Program: Increasing Science Outcomes for All Students. *Remedial and Special Education: RASE*, *42*(3), 140–154. https://doi.org/10.1177/0741932521989091

Durmuş, J., & Bayraktar, Ş. (2010). Effects of Conceptual Change Texts and Laboratory Experiments on Fourth Grade Students' Understanding of Matter and Change Concepts. *Journal of Science Education and Technology*, *19*(5), 498–504. https://link.springer.com/article/10.1007/s10956-010-9216-9

Eysink, T. H. S., Hulsbeek, M., & Gijlers, H. (2017). Supporting primary school teachers in differentiating in the regular classroom. *Teaching and Teacher Education*, *66*, 107–116. https://doi.org/10.1016/j.tate.2017.04.002

Fasasi, R. A. (2017). Effects of ethnoscience instruction, school location, and parental educational status on learners' attitude towards science. *International Journal of Science Education*, 39(5), 548-564. https://doi.org/10.1080/09500693.2017.1296599

Ferrell, K. A., Correa-Torres, S. M., Howell, J. J., Pearson, R., Carver, W. M., Groll, A. S., Anthony, T. L., Matthews, D., Gould, B., O'Connell, T., Botsford, K. D., Dewald, H. P., Smyth, C. A., & Dewald, A. J. (2017). Audible Image Description as an Accommodation in Statewide Assessments for Students with Visual and Print Disabilities. *Journal of Visual Impairment & Blindness*, *111*(4), 325–339. https://doi.org/10.1177/0145482X1711100403

Glick, M. P., & Samarapungavan, A. (2008). Wolves are Beautiful and Proud. *Journal of Museum Education*, *33*(2), 199–207. https://doi.org/10.1080/10598650.2008.11510600

Goldschmidt, P., & Jung, H. (2011). *Evaluation of Seeds of Science/Roots of Reading: effective tools for developing literacy through science in the early grades Light Energy unit* (No. 781). National Center for Research on Evaluation, Standards, and Student Testing (CRESST). https://files.eric.ed.gov/fulltext/ED520425.pdf

Hand, B., Shelley, M. C., Laugerman, M., Fostvedt, L., & Therrien, W. (2018). Improving critical thinking growth for disadvantaged groups within elementary school science: A randomized controlled trial using the Science Writing Heuristic approach. *Science Education*, *102*(4), 693–710. https://doi.org/10.1002/sce.21341

Hanley, P., Slavin, R., & Elliott, L. (2015). *Thinking, Doing, Talking Science: Evaluation report and executive summary*. EEF. https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation_Reports/Oxford_Science.pdf

Hanley, P., Wilson, H., Holligan, B., & Elliott, L. (2020). Thinking, doing, talking science: the effect on attainment and attitudes of a professional development programme to provide cognitively challenging primary science lessons. *International Journal of Science Education*, *42*(15), 2554–2573. https://doi.org/10.1080/09500693.2020.1821931

Hardiman, M. M., JohnBull, R. M., Carran, D. T., & Shelton, A. (2019). The effects of arts-integrated instruction on memory for science content. *Trends in Neuroscience and Education*, *14*, 25–32. https://doi.org/10.1016/j.tine.2019.02.002

Henrichs, L. F., & Leseman, P. P. M. (2014). Early Science Instruction and Academic Language Development Can Go Hand in Hand. The Promising Effects of a Low-Intensity Teacher-Focused Intervention. *International Journal of Science Education*, *36*(17), 2978–2995. https://doi.org/10.1080/09500693.2014.948944

Hodges, G. W., Flanagan, K., Lee, J., Cohen, A., Krishnan, S., & Ward, C. (2020). A quasi-experimental study comparing learning gains associated with serious educational gameplay and hands-on science in elementary classrooms. *Journal of Research in Science Teaching*, *57*(9), 1460–1489. https://doi.org/10.1002/tea.21661

Hondrich, A. L., Decristan, J., Hertel, S., & Klieme, E. (2018). Formative assessment and intrinsic motivation: The mediating role of perceived competence. *Zeitschrift Für Erziehungswissenschaft*, *21*(4), 717–734. https://link.springer.com/article/10.1007/s11618-018-0833-z

Hu, H.-W., Chiu, C.-H., & Chiou, G.-F. (2019). Effects of question stem on pupils' online questioning, science learning, and critical thinking. *The Journal of Educational Research*, *112*(4), 564–573. https://doi.org/10.1080/00220671.2019.1608896

Hwang, G.-J., & Chang, S.-C. (2021a). Facilitating knowledge construction in mobile learning contexts: A bi-directional peer-assessment approach. *British Journal of Educational Technology: Journal of the Council for Educational Technology*, *52*(1), 337–357.

Hwang, G.-J., Chang, S.-C., Song, Y., & Hsieh, M.-C. (2021b). Powering up flipped learning: An online learning environment with a concept map-guided problem-posing strategy. *Journal of Computer Assisted Learning*, *37*(2), 429–445. https://doi.org/10.1111/jcal.12499

Hwang, G.-J., Tu, N.-T., & Wang, X.-M. (2018). The Impacts of Guided Peer-Feedback on Students' Learning Achievements and Project Outcomes in Science Courses. *Journal of Educational Technology & Society*, *21*(1), 25–36. https://www.jstor.org/stable/26273865

Jay, T., Willis, B., Thomas, P., Taylor, R., Moore, N., Burnett, C., Merchant, G., & Stevens, A. (2017). *Dialogic Teaching: Evaluation Report and Executive Summary*. Education Endowment Foundation (EEF). http://shura.shu.ac.uk/17014/1/Dialogic_Teaching_Evaluation_Report.pdf

Kara, S., & Kingir, S. (2022). Implementation of the Model-Based Science Writing Heuristic Approach in Elementary School Science. *International Journal of Science and Mathematics Education*, *20*(4), 683–703. https://link.springer.com/article/10.1007/s10763-021-10191-0

Kim, J. S., Relyea, J. E., Burkhauser, M. A., Scherer, E., & Rich, P. (2021). Improving Elementary Grade Students' Science and Social Studies Vocabulary Knowledge Depth, Reading Comprehension, and Argumentative Writing: a Conceptual Replication. *Educational Psychology Review*, *33*(4), 1935–1964. https://link.springer.com/article/10.1007/s10648-021-09609-6

Kim, K. H., VanTassel-Baska, J., Bracken, B. A., Feng, A., Stambaugh, T., & Bland, L. (2012). Project Clarion: Three Years of Science Instruction in Title I Schools among K-Third Grade Students. *Research in Science Education*, *42*(5), 813–829. https://link.springer.com/article/10.1007/s11165-011-9218-5

Kim, P., & Olaciregui, C. (2008). The effects of a concept map-based information display in an electronic portfolio system on information processing and retention in a fifth-grade science class covering the Earth's atmosphere. *British Journal of Educational Technology: Journal of the Council for Educational Technology*, *39*(4), 700–714. https://doi.org/10.1111/j.1467-8535.2007.00763.x

Kitmitto, S., González, R., Mezzanote, J., & Chen, Y. (2018). *Thinking, Doing, Talking Science: Evaluation report and executive summary.* Education Endowment Foundation (EEF). https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation_Reports/TDTS.pdf

Lai, C.-L., Hwang, G.-J., & Tu, Y.-H. (2018). The effects of computer-supported self-regulation in science inquiry on learning outcomes, learning processes, and self-efficacy. *Educational Technology Research and Development: ETR & D*, *66*(4), 863–892. https://link.springer.com/article/10.1007/s11423-018-9585-y

Lai, C.-S. (2016). Third Graders' Understanding of Air Concepts Facilitated by the iPod Inquiry Teaching Method. *International Journal of Research in Education and Science*, *2*(1), 1–9. https://files.eric.ed.gov/fulltext/EJ1105183.pdf

Lai, C.-S., & Chan, K.-L. (2020). Enhancing Science Learning through Science Trade Book Reading for 5th Graders. *Journal of Education in Science, Environment and Health*, *6*(1), 1–9. https://doi.org/10.21891/jeseh.669294

Letina, A. (2016). Effectiveness of Inquiry-Based Science and Social Studies Teaching in the Development of Students' Scientific Competence. *Croatian Journal of Education: Hrvatski časopis Za Odgoj I Obrazovanje*, *18*(3), 665–696. https://doi.org/10.15516/cje.v18i3.1735

Letina, A. (2020). Development of Students' Learning to Learn Competence in Primary Science. *Education Sciences*, *10*(11), 325. https://doi.org/10.3390/educsci10110325

Leuchter, M., Saalbach, H., & Hardy, I. (2014). Designing Science Learning in the First Years of Schooling. An intervention study with sequenced learning material on the topic of "floating and sinking." *International Journal of Science Education*, *36*(10), 1751–1771. https://doi.org/10.1080/09500693.2013.878482

Li, Y., Huang, Z., Jiang, M., & Chang, T.-W. (2016). The Effect on Pupils' Science Performance and Problem-Solving Ability through Lego: An Engineering Design-based Modeling Approach. *Journal of Educational Technology & Society*, *19*(3), 143–156. https://www.jstor.org/stable/10.2307/jeductechsoci.19.3.143

Lin, H., Hong, Z., & Cheng, Y. (2009). The Interplay of the Classroom Learning Environment and Inquiry-based Activities. *International Journal of Science Education*, *31*(8), 1013–1024. https://doi.org/10.1080/09500690701799391

Llosa, L., Lee, O., Jiang, F., Haas, A., O'Connor, C., Van Booven, C. D., & Kieffer, M. J. (2016). Impact of a Large-Scale Science Intervention Focused on English Language Learners. *American Educational Research Journal*, *53*(2), 395–424. https://journals.sagepub.com/doi/pdf/10.3102/0002831216637348?

Looi, C.-K., Chen, W., & Ng, F.-K. (2010). Collaborative activities enabled by GroupScribbles (GS): An exploratory study of learning effectiveness. *Computers & Education*, *54*(1), 14–26. https://doi.org/10.1016/j.compedu.2009.07.003

Looi, C.-K., Zhang, B., Chen, W., Seow, P., Chia, G., Norris, C., & Soloway, E. (2011). 1:1 mobile inquiry learning experience for primary science students: a study of learning effectiveness. *Journal of Computer Assisted Learning*, *27*(3), 269–287. https://doi.org/10.1111/j.1365-2729.2010.00390.x

Maerten-Rivera, J., Ahn, S., Lanier, K., Diaz, J., & Lee, O. (2016). Effect of a Multiyear Intervention on Science Achievement of All Students including English Language Learners. *The Elementary School Journal*, *116*(4), 600–624.

Meyer, A., Klingenberg, K., & Wilde, M. (2016). The Benefits of Mouse Keeping—an Empirical Study on Students' Flow and Intrinsic Motivation in Biology Lessons. *Research in Science Education*, *46*(1), 79–90. https://link.springer.com/article/10.1007/s11165-014-9455-5

Michalsky, T., Mevarech, Z. R., & Haibi, L. (2009). Elementary School Children Reading Scientific Texts: Effects of Metacognitive Instruction. *The Journal of Educational Research*, *102*(5), 363–376.

Miller, B. W., Anderson, R. C., Morris, J., Lin, T.-J., Jadallah, M., & Sun, J. (2014). The effects of reading to prepare for argumentative discussion on cognitive engagement and conceptual growth. *Learning and Instruction*, *33*, 67–80. 10.1016/j.learninstruc.2014.04.003

Mills, L. A., & Katzman, W. (2015). Examining the Effects of Field Trips on Science Identity. International Association for Development of the Information Society. Paper presented at the International Association for Development of the Information Society (IADIS) International Conference on Cognition and Exploratory Learning in the Digital Age (CELDA) (12th, Maynooth, Greater Dublin, Ireland, Oct 24-26, 2015)

Olgun, O. S., & Adali, B. (2008). Teaching grade 5 life science with a case study approach. *Journal of Elementary Science Education*, *20*(1), 29–44. https://link.springer.com/article/10.1007/BF03174701

Piila, E., Salmi, H., & Thuneberg, H. (2021). STEAM-learning to Mars: students' ideas of space research. *Education Sciences*, *11*(3), 122. https://doi.org/10.3390/educsci11030122

Polikoff, M., Le, Q. T., Danielson, R. W., Sinatra, G. M., & Marsh, J. A. (2018). The Impact of Speedometry on Student Knowledge, Interest, and Emotions. *Journal of Research on Educational Effectiveness*, *11*(2), 217–239. https://doi.org/10.1080/19345747.2017.1390025

Qiao, X., & Zhou, X. (2020). Research on the Integration of STEM Education into the Rural Elementary School Science Curriculum: An Example from Rural Elementary Schools in Western China. *Best Evidence in Chinese Education*, *5*(1), 581–590. https://dx.doi.org/10.2139/ssrn.3607636

Randler, C. (2009). Learning About Bird Species on the Primary Level. *Journal of Science Education and Technology*, *18*(2), 138–145. https://link.springer.com/article/10.1007/s10956-008-9139-x

Reeves, C. G., Fostvedt, L., Laugerman, M., Baenziger, J., Shelley, M., Hand, B., & Therrien, W. (2013). Structural Equation Modeling of Knowledge Content Improvement Using Inquiry Based Instruction. *Society for Research on Educational Effectiveness*. http://files.eric.ed.gov/fulltext/ED564105.pdf

Rotgans, J. I., & Schmidt, H. G. (2017). Interest development: Arousing situational interest affects the growth trajectory of individual interest. *Contemporary Educational Psychology*, *49*, 175–184. https://doi.org/10.1016/j.cedpsych.2017.02.003

Schalk, L., Edelsbrunner, P. A., Deiglmayr, A., Schumacher, R., & Stern, E. (2019). Improved application of the control-of-variables strategy as a collateral benefit of inquiry-based physics education in elementary school. *Learning and Instruction*, *59*, 34–45. https://doi.org/10.1016/j.learninstruc.2018.09.006

Scott, G. W., & Boyd, M. (2016). Getting more from getting out: increasing achievement in literacy and science through ecological fieldwork. *Education 3-13*, *44*(6), 661–670. http://dx.doi.org/10.1080/03004279.2014.996242

Sternberg, R. J., Jarvin, L., Birney, D. P., Naples, A., Stemler, S. E., Newman, T., Otterbach, R., Parish, C., Randi, J., & Grigorenko, E. L. (2014). Testing the theory of successful intelligence in teaching grade 4 language arts, mathematics, and science. *Journal of Educational Psychology*, *106*(3), 881–899.

Sun, K.-T., Lin, C.-L., & Wang, S.-M. (2010). A 3-D virtual reality model of the sun and the moon for e-learning at elementary schools. *International Journal of Science and Mathematics Education*, *8*(4), 689–710. https://link.springer.com/article/10.1007/s10763-009-9181-z

Tsai, C.-Y., Jack, B. M., Huang, T.-C., & Yang, J.-T. (2012). Using the Cognitive Apprenticeship Web-based Argumentation System to Improve Argumentation Instruction. *Journal of Science Education and Technology*, *21*(4), 476–486. https://link.springer.com/article/10.1007/s10956-011-9339-7

Ünal, M., & Aral, N. (2014). An investigation on the effects of experiment based education program on six years olds' problem solving skills. *Education and Science - EĞİTİM VE BİLİM*, *39*(176), 279–291.

Upadhyay, B., & DeFranco, C. (2008). Elementary students' retention of environmental science knowledge: Connected science instruction versus direct instruction. *Journal of Elementary Science Education*, *20*(2), 23–37.

van der Graaf, J., van de Sande, E., Gijsel, M., & Segers, E. (2019). A combined approach to strengthen children's scientific thinking: direct instruction on scientific reasoning and training of teacher's verbal support. *International Journal of Science Education*, *41*(9), 1119–1138. https://doi.org/10.1080/09500693.2019.1594442

Vitale, M. R., & Romance, N. R. (2012). Using in-depth science instruction to accelerate student achievement in science and reading comprehension in grades 1-2. *International Journal of Science and Mathematics Education*, *10*(2), 457–472. https://link.springer.com/article/10.1007/s10763-011-9326-8

Wang, T.-L., & Tseng, Y.-K. (2018). The Comparative Effectiveness of Physical, Virtual, and Virtual-Physical Manipulatives on Third-Grade Students' Science Achievement and Conceptual Understanding of Evaporation and Condensation. *International Journal of Science and Mathematics Education*, *16*(2), 203–219. https://link.springer.com/article/10.1007/s10763-016-9774-2

Ward, W., Cole, R., Bolaños, D., Buchenroth-Martin, C., Svirsky, E., & Weston, T. (2013). My science tutor: A conversational multimedia virtual tutor. *Journal of Educational Psychology*, *105*(4), 1115–1125.

Wells, N. M., Myers, B. M., Todd, L. E., Barale, K., Gaolach, B., Ferenz, G., Aitken, M., Henderson, C. R., Tse, C., Pattison, K. O., Taylor, C., Connerly, L., Carson, J. B., Gensemer, A. Z., Franz, N. K., & Falk, E. (2015). The Effects of School Gardens on Children's Science Knowledge: A randomized controlled trial of low-income elementary schools. *International Journal of Science Education*, *37*(17), 2858–2878. https://doi.org/10.1080/09500693.2015.1112048

Williams, J. P., Stafford, K. B., Lauer, K. D., Hall, K. M., & Pollini, S. (2009). Embedding reading comprehension training in content-area instruction. *Journal of Educational Psychology*, *101*(1), 1–20.

Wünschmann, S., Wüst-Ackermann, P., Randler, C., Vollmer, C., & Itzek-Greulich, H. (2017). Learning Achievement and Motivation in an Out-of-School Setting—Visiting Amphibians and Reptiles in a Zoo Is More Effective than a Lesson at School. *Research in Science Education*, *47*(3), 497–518. https://link.springer.com/article/10.1007/s11165-016-9513-2

Yeo, J., Lim, E., Tan, K. C. D., & Ong, Y. S. (2021). The Efficacy of an Image-to-Writing Approach to Learning Abstract Scientific Concepts: Temperature and Heat. *International Journal of Science and Mathematics Education*, *19*(1), 21–44. https://link.springer.com/article/10.1007/s10763-019-10026-z

Zacharia, Z. C., Lazaridou, C., & Avraamidou, L. (2016). The use of mobile devices as means of data collection in supporting elementary school students' conceptual understanding about plants. *International Journal of Science Education*, *38*(4), 596–620. https://doi.org/10.1080/09500693.2016.1153811

Zangori, L., Forbes, C. T., & Schwarz, C. V. (2015). Exploring the Effect of Embedded Scaffolding Within Curricular Tasks on Third-Grade Students' Model-Based Explanations about Hydrologic Cycling. *Science & Education*, *24*(7), 957–981.

Zhang, D., & Campbell, T. (2012). An Exploration of the Potential Impact of the Integrated Experiential Learning Curriculum in Beijing, China. *International Journal of Science Education*, *34*(7), 1093–1123.

Zhang, L. (2018). Withholding answers during hands-on scientific investigations? Comparing effects on developing students' scientific knowledge, reasoning, and application. *International Journal of Science Education*, *40*(4), 459–469.

Zheng, R. Z., Yang, W., Garcia, D., & McCadden, E. P. (2008). Effects of multimedia and schema induced analogical reasoning on science learning. *Journal of Computer Assisted Learning*, *24*(6), 474–482.

Zwiep, S. G., & Straits, W. J. (2013). Inquiry Science: The Gateway to English Language Proficiency. *Journal of Science Teacher Education*, *24*(8), 1315–1331.

References

Abrahams, I.Z & Reiss, M.J. (2016) *Effective Practical Science 11-16*. London: Bloomsbury.

Abrahams, I., Reiss, M., & Sharpe, R. (2014). 'The impact of "Getting practical work in science" continuing professional development programme on teachers' ideas and practice in science practical work', *Research in Science and Technological Education*, 32 (3), 263–280.

Adey, P. (2005). Issues arising from the long-term evaluation of cognitive acceleration programs. *Research in Science Education*, *35*, 3-22.

Banerjee, P.A. (2016) A systematic review of factors linked to poor academic performance of disadvantaged students in science and maths in schools. *Cogent Education*, 3 (1), 2-17. 10.1080/2331186X.2016.1178441.

Bennett, J., Hanley, P., Abrahams, I., Elliott, L., & Turkenburg-van Diepen, M. (2019) Mixed methods, mixed outcomes? Combining an RCT and case studies to research the impact of a training programme for primary school science teachers. *International Journal of Science Education* 41(4), 490–509. Routledge.

Bennett, J., Lubben, F., & Hogarth, S. (2007). Bringing science to life: A synthesis of the research evidence on the effects of context-based and STS approaches to science teaching. *Science education*, *91*(3), 347-370. https://doi.org/10.1002/sce.20186

Bianchi, L., Whittaker, C. and Poole, A. (2021) The 10 key issues with children's learning in primary science in England. The University of Manchester and The Ogden Trust. (https://seerih-innovations.org/just-good-stuff/10-key-issues)

Black, P., & Wiliam, D. (2003). 'In praise of educational research': Formative assessment. *British Educational Research Journal*, *29*(5), 623-637.

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2004). Working inside the black box: Assessment for Learning in the Classroom. *Phi Delta Kappan*, 86(1), 8–21.

Bloom, B. S. (1956) Taxonomy of Educational Objectives, Handbook I: The Cognitive Domain. New York: David McKay Co Inc.

Boyle, B., & Bragg, J. (2005). No science today—the demise of primary science. *The Curriculum Journal*, *16*(4), 423-437. https://doi.org/10.1080/09585170500384115

Braund, M. & Reiss, M. (2004). *Learning science outside the classroom.* London: Routledge

Bybee, R. (1997). Achieving scientific literacy: From purposes to practical action. Portsmouth, NH: Heinemann.

Chang, Y. H., Chang, C. Y., & Tseng, Y. H. (2010). Trends of science education research: An automatic content analysis. *Journal of Science Education and Technology*, *19*(4), 315-331. https://link.springer.com/article/10.1007/s10956-009-9202-2

Child Welfare League Foundation (2022). Foundation warns of widening education gap. Taipai Times Thursday, February 3rd, 2022 page2. Available at: https://www.taipeitimes.com/News/taiwan/archives/2022/02/03/2003772516

Chinn, S. (2000). A simple method for converting an odds ratio to effect size for use in meta-analysis. *Statistics in Medicine*, 19(22), 3127-3131. https://doi.org/10.1002/1097-0258(20001130)19:22<3127::AID-SIM784>3.0.CO;2-M

Chou, R. (2008). Using evidence in pain practice: Part I: Assessing quality of systematic reviews and clinical practice guidelines. *Pain Medicine*, 9(5), 518-530. https://doi.org/10.1111/j.1526-4637.2008.00422_1.x

Collins, S., Reiss, M., & Stobart, G. (2010). What happens when high-stakes testing stops? Teachers' perceptions of the impact of compulsory national testing in science of 11-year-olds in England and its abolition in Wales. *Assessment in Education: Principles, Policy & Practice*, *17*(3), 273-286. https://doi.org/10.1080/0969594X.2010.496205

Cobern, W.W., Schuster, D., Adams, B., Applegate, B., Skjold, B., Undreiu, A., Loving, C.C. & Janice D. Gobert, J.D. (2010) Experimental comparison of inquiry and direct instruction in science, *Research in Science & Technological Education*, 28 (1), 81-96.

Cohen, J. (1992) Quantitative Methods in Psychology: a Power Primer. *Psychological Bulletin*, 112 (1), 155-159.

Department for Education (2021). *Research review series: science.* https://www.gov.uk/government/publications/research-review-series-science/research-review-series-science

Department of Education and Employment, DfEE (1988) The National Literacy Strategy. London: DfEE.

Department of Education and Employment, DfEE (1989) The National Numeracy Strategy: London: DfEE.

Department for Education and Science (1989). *The Education (National Curriculum) (Attainment Targets and Programmes of Study in English).* DfES.

Department for Education and Employment (1999). *Introducing the Revised National Curriculum.* London: DfEE/QCA. Available at: https://webarchive.nationalarchives.gov.uk/ukgwa/20130401151715/https:/www.education.gov.uk/publications/eOrderingDownload/QCA-99-457.pdf

Department for Education (2013). The national curriculum in England: key stages 1 and 2 framework document. Available at: https://www.gov.uk/government/publications/national-curriculum-in-england-primary-curriculum

DES and WO (1985) Science 5-16: A Statement of Policy. London, HMSO.

DeWitt, J. & Archer, L. (2015) Who Aspires to a Science Career? A comparison of survey responses from primary and secondary school students, *International Journal of Science Education*, 37 (13), 2170-2192.https://doi.org/10.1080/09500693.2015.1071899.

DeWitt, J., Archer, L., & Osborne, J. (2014). Science-related aspirations across the primary–secondary divide: Evidence from two surveys in England. *International Journal of Science Education*, *36*(10), 1609-1629. https://doi.org/10.1080/09500693.2013.871659.

Dillenbourg, P. (1999). What do you mean by ‘collaborative learning’? In P. Dillenbourg (Ed.), Collaborative learning (pp. 1–19). Oxford, England: Routledge.

Dowd, J. E., Thompson Jr, R. J., Schiff, L. A., & Reynolds, J. A. (2018). Understanding the complex relationship between critical thinking and science reasoning among undergraduate thesis writers. *CBE—Life Sciences Education*, *17*(1), ar4. https://doi.org/10.1187/cbe.17-03-0052

Driver, R. (1983) *The pupil as scientist?* Buckingham: Open University Press

Earle, S. (2014). Formative and summative assessment of science in English primary schools: evidence from the Primary Science Quality Mark. *Research in Science & Technological Education*, *32*(2), 216-228. https://doi.org/10.1080/02635143.2014.913129

Earle, S. (2019). *Assessment in the Primary Classroom: Principles and practice*. Learning Matters. Sage London.

Earle, S., & Turner, J. (2023). What has happened to teacher assessment of science in English primary schools? Revisiting evidence from the Primary Science Quality Mark. *Research in Science & Technological Education*, *41*(1), 22-38. https://doi.org/10.1080/02635143.2020.1834375

Education Endowment Foundation. (2018). Improving Secondary Science: Guidance Report. Education Endowment Foundation. https://files.eric.ed.gov/fulltext/ED612065.pdf

Education Endowment Foundation. (2021). Systematic review of evidence on primary science teaching: call for proposals. https://educationendowmentfoundation.org.uk/public/files/ITT_-_Systematic_review_of_evidence_on_primary_science_teaching.pdf

Education Endowment Foundation. (2021). Year 6 science assessment https://educationendowmentfoundation.org.uk/projects-and-evaluation/evaluation/eef-outcome-measures-and-databases/year-6-science-assessment

EPHPP (Effective Public Health Practice Project) (n.d.) Quality Assessment tool for Quantitative Studies. https://merst.healthsci.mcmaster.ca/ephpp/

Erduran, S., Simon, S., & Osborne, J. (2004). TAPping into argumentation: Developments in the application of Toulmin's argument pattern for studying science discourse. *Science Education*, 88(6), 915-933. https://onlinelibrary.wiley.com/doi/abs/10.1002/sce.20012

Fuller, R., Landrigan, P. J., Balakrishnan, K., Bathan, G., Bose-O'Reilly, S., Brauer, M., ... & Yan, C. (2022). Pollution and health: a progress update. *The Lancet Planetary Health*, *6*(6), e535-e547. https://doi.org/10.1016/S2542-5196(22)00090-0

Gilbert, J. K., Bulte, A. M. W., & Pilot, A. (2011). Concept Development and Transfer in Context-Based Science Education. International Journal of Science Education, 33(6), 817–837. https://doi.org/10.1080/09500693.2010.493185

Gough, D., Oliver, S. & Thomas, J. (2017). Introducing systematic reviews. In D. Gough, S. Oliver & J. Thomas (Eds.), An introduction to systematic reviews (2nd edition, pp. 1–18). London: Sage.

Grangeat, M., Harrison, C., & Dolin, J. (2021). Exploring assessment in STEM inquiry learning classrooms. *International Journal of Science Education*, *43*(3), 345-361. https://doi.org/10.1080/09500693.2021.1903617

Green, F. (2020). *Schoolwork in lockdown: new evidence on the epidemic of educational poverty*. Centre for Learning and Life Chances in Knowledge Economies and Societies. Available at: https://www.llakes.ac.uk/wp-content/uploads/2021/03/RP-67-Francis-Green-Research-Paper-combined-file.pdf

Gresnigt, R. Taconis, R., van Keulen, H., Gravemeijer, K & Baartman, L. (2014) Promoting science and technology in primary education: a review of integrated curricula, *Studies in Science Education*, 50:1,47-84, DOI: 10.1080/03057267.2013.877694.

Guerra-Ramos, M. T., Ryder, J., & Leach, J. (2010). Ideas about the nature of science in pedagogically relevant contexts: Insights from a situated perspective of primary teachers' knowledge. *Science Education*, *94*(2), 282-307. https://doi.org/10.1002/sce.20361

Haidich, A. B. (2010). Meta-analysis in medical research. *Hippokratia*, 14(Suppl 1), 29.

Harlen, W. (2007). Assessment of learning. Sage.

Harlen, W., Bell, D., Devés, R., Dyasi, H., Fernández de la Garza, G., Léna, P., Millar, R., Reiss, M., Rowell, P., & Yu, W. (2010). Principles and big ideas of science education. Hatfield: Association for Science Education. https://www.ase.org.uk/bigideas

Harlen, W. (2013). Inquiry-based learning in science and mathematics. *Review of science, mathematics and ICT education*, 7(2), 9-33.

Harlen, W. (2018) The teaching of science in primary schools (Seventh edition). London and New York: Routledge.

Hartmeyer, R., Stevenson, M.P., & Bentsen, P. (2018). A systematic review of concept mapping-based formative assessment processes in primary and secondary science education. *Assessment in Education: Principles, Policy & Practice* 25(6), 598–619. Routledge.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112.

Huerta, M., Garza, T. (2019) Writing in Science: Why, How, and for Whom? A Systematic Literature Review of 20 Years of Intervention Research (1996–2016). E*ducational Psychology Review,* 31, 533–570. https://doi.org/10.1007/s10648-019-09477-1.

Hughes, C. A., Morris, J. R., Therrien, W. J., & Benson, S. K. (2017). Explicit instruction: Historical and contemporary contexts. *Learning Disabilities Research & Practice*, *32*(3), 140-148. https://doi.org/10.1111/ldrp.12142

Hutchings, M. (2016). Exam factories? The impact of accountability measures on children and young people. https://www.basw.co.uk/system/files/resources/basw_112157-4_0.pdf

IPCC (2023) AR6 Synthesis Report: Climate Change 2023. https://www.ipcc.ch/report/sixth-assessment-report-cycle/

James, M., & Brown, S. (2007). Grasping the TLRP nettle: preliminary analysis and some enduring issues surrounding the improvement of learning outcomes. *The Curriculum Journal*, 16(1), 7-30. https://www.tandfonline.com/doi/full/10.1080/0958517042000336782

Keys, C. W., Hand, B., Prain, V., & Collins, S. (1999). Using the science writing heuristic as a tool for learning from laboratory investigations in secondary science. *Journal of Research in Science Teaching*, 36(10), 1065-1084.

Kuhn, J., & Müller, A. (2014). Context-based science education by newspaper story problems: A study on motivation and learning effects. *Perspectives in Science*, *2*(1-4), 5-21. https://doi.org/10.1016/j.pisc.2014.06.001

Lee, H., Chung, H. Q., Zhang, Y., Abedi, J., & Warschauer, M. (2020). The effectiveness and features of formative assessment in US K-12 education: A systematic review. *Applied Measurement in Education*, 33(2), 124-140.

Lee, M. H., Wu, Y. T., & Tsai, C. C. (2009). Research trends in science education from 2003 to 2007: A content analysis of publications in selected journals. *International Journal of Science Education*, *31*(15), 1999-2020. https://doi.org/10.1080/09500690802314876

Lelliott, A. & Rollnick, M. (2010) Big Ideas: A review of astronomy education research 1974–2008, *International Journal of Science Education*, 32 (13), 1771-1799. https://doi.org/10.1080/09500690903214546.

Leonardi, S., Lamb, H., Howe, P., Choudhury, A., & (CFE Research). (2017). *"State of the nation" report of UK primary science education: baseline research for the Wellcome Trust Primary Science Campaign*. Wellcome Trust. https://wellcome.ac.uk/sites/default/files/state-of-the-nation-report-of-uk-science-education.pdf

Mackintosh, J., White, E., & Dickerson, C. (2017). Developing teachers as leaders of science in primary schools. *Journal of Emergent Science*. Association for Science Education (ASE). http://uhra.herts.ac.uk/handle/2299/19062

McCrory, A. (2017). Scientific enquiry and engaging primary-aged children in science lessons: reflections on provision for scientific enquiry in primary schools in England and Wales: Part 1. *Journal of Emergent Science* 13. https://discovery.ucl.ac.uk/id/eprint/1574818/.

McGarr, O. & Johnston, K. (2020). Curricular responses to Computer Science provision in schools: current provision and alternative possibilities. *Curriculum Journal,* Vol 31, (4), pp.745-756.

McLeod, L. (2018). Developing a framework for the biology curriculum. *School Science Review* 100(370), 23-29.

Millar, R., & Osborne, J. (1998). Beyond 2000. Science education for the future. London: Nuffield Foundation.

Millar, R. (2006). Twenty First Century Science: Insights from the Design and Implementation of a Scientific Literacy Approach in School Science. International Journal of Science Education 28(13),1499–1521.

Mourshed, M., Krawitz, M., & Dorn, E. (2017). How to improve student educational outcomes: New insights from data analytics. McKinsey & Company.

Mullis, I.V.S., Martin, M.O., Foy, P., Kelly, D.L. & Bethany Fishbein, B. (2020) TIMSS 2019 International Results in Mathematics and Science. Boston: Lynch School of Education and Human Development. Retrieved from Boston College, TIMSS &

PIRLS International Study Center website: https://timssandpirls.bc.edu/timss2019/international-results/.

Murphy, C. and Beggs, J. 2003a. Primary pupils' and teachers' use of computers at home and school. *British Journal of Educational Technology*, 34(1), 79–83.

Murphy, C., Neil, P., & Beggs, J. (2007). Primary science teacher confidence revisited: Ten years on. *Educational research*, *49*(4), 415-430. https://doi.org/10.1080/00131880701717289

National Center on Education and the Economy (NCEE) (2022). Taiwan, Digital Platforms and Resources. Available at: https://ncee.org/country/taiwan/

Natural History Museum (2023). National Education Nature Park and Climate Action Awards https://www.nhm.ac.uk/about-us/national-impact/national-education-nature-park-and-climate-action-awards-scheme.html

Next Generation Science Standards NGSS (n.d.) The three dimensions of science learning. https://www.nextgenscience.org/

Norris, S., & Phillips, L. (2003). How literacy in its fundamental sense is central to scientific literacy. Science Education, 87(2),224–240.

Nunes, T., Bryant, P., Strand, S., Hillier, J., Barros, R. & Miller-Friedmann, J. (2017). Review of SES and Science Learning in Formal Educational Settings. EEF and Royal Society. https://royalsociety.org/~/media/policy/topics/education-skills/education-research/evidence-review-eef-royalsociety-22-09-2017.pdf?la=en-GB

Ofsted (2021). Research review series: science. https://www.gov.uk/government/publications/research-review-series-science/research-review-series-science

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

OECD (2016), PISA 2015 Results (Volume II): Policies and Practices for Successful Schools, PISA, OECD Publishing, Paris. http://dx.doi.org/10.1787/9789264267510-en

Osborne, J.F., Henderson, J.B., MacPherson, A., Szu, E., Wild, A. and Yao, S.Y., (2016). The development and validation of a learning progression for argumentation in science. *Journal of Research in Science Teaching*, 53(6), 821-846. https://onlinelibrary.wiley.com/doi/full/10.1002/tea.21316

Popay, J., Roberts, H., Sowden, A., Petticrew, M., Arai, L., Rodgers, M., ... & Duffy, S. (2006). Guidance on the conduct of narrative synthesis in systematic reviews. *A product from the ESRC methods programme Version*, *1*(1), b92.

Parliamentary Office for Science and Technology (POST). 2003. *Primary science*, London: POST. https://post.parliament.uk/research-briefings/post-pn-202/

Podolsky, A., Kini, T., & Darling-Hammond, L. (2019). Does teaching experience increase teacher effectiveness? A review of US research. *Journal of Professional Capital and Community*, 4(4), 286-308.

Potvin, P. & Hasni, A. (2014) Interest, motivation and attitude towards science and technology at K-12 levels: a systematic review of 12 years of educational research, *Studies in Science Education*, 50 (1), 85-129.

Ramsden, J.M. (1997) How does a context-based approach influence understanding of key chemical ideas at 16+? *International Journal of Science Education*, 19 (6), 697-710.

Rosenshine, B. (2012). Principles of instruction: Research-based strategies that all teachers should know. *American educator*, *36*(1), 12. https://files.eric.ed.gov/fulltext/EJ971753.pdf

Royal Society. (2010). Primary science and mathematics education: getting the basics right. Available from: https://royalsociety.org/-/media/Royal_Society_Content/Education/policy/state-of-nation/2010-07-07-SNR3-Basics.pdf

Rushton, E.A.C. & Reiss, M.J. (2021) Middle and high school science teacher identity considered through the lens of the social identity approach: a systematic review of the literature. *Studies in Science Education*, 57 (2), 141-203.

Samson, D., & Schoelles, K. M. (2012). Developing the topic and structuring systematic reviews of medical tests: utility of PICOTS, analytic frameworks, decision trees, and other frameworks. *Methods Guide for Medical Test Reviews*. In: Chang SM, Matchar DB, Smetana GW, *et al.*, editors. Methods Guide for Medical Test Reviews [Internet]. Rockville (MD): Agency for Healthcare Research and Quality (US); 2012 Jun. Chapter 2. Available from: https://www.ncbi.nlm.nih.gov/sites/books/NBK98235/.

Schoor, C., Narciss, S., & Körndle, H. (2015). Regulation during cooperative and collaborative learning: A theory-based review of terms and concepts. *Educational Psychologist*, 50(2), 97-119.

Standards and Testing Agency, STA (2019) Key stage 2 science sampling 2018. London: The Standards and Testing Agency.

Sheskin, D. (2004). Handbook of parametric and nonparametric statistical procedures (3rd ed.). Boca Raton, FL: Chapman & Hall/CRC.

Slavin, R.E. (1980) Co-operative Learning. *Review of Educational Research*, 50 (2), 315-342.

Slavin, R. E. (1987). Mastery learning reconsidered. *Review of educational research*, *57*(2), 175-213. https://journals.sagepub.com/doi/abs/10.3102/00346543057002175

Slavin, R.E., Lake, C., Hanley, P. and Thurston, A. (2014) Experimental evaluations of elementary science programs: A best-evidence synthesis. *Journal of Research in Science Teaching*, 51 (7), 870-901. https://doi.org/10.1002/tea.21139

Tolmie, A. K., Topping, K. J., Christie, D., Donaldson, C., Howe, C., Jessiman, E., ... & Thurston, A. (2010). Social effects of collaborative learning in primary schools. *Learning and instruction*, *20*(3), 177-191. https://doi.org/10.1016/j.learninstruc.2009.01.005

Trickey, S., & Topping, K. J. (2004). ‘Philosophy for children’: A systematic review. *Research papers in Education*, 19(3), 365-380.

UK Parliament. (2017). *Primary assessment*
https://publications.parliament.uk/pa/cm201617/cmselect/cmeduc/682/68202.htm

UNEP, Convention on Biological Diversity and WHO. (2015). Connecting Global Priorities: Biodiversity and Human Health A State of Knowledge Review. https://iris.who.int/bitstream/handle/10665/174012/9789241508537_eng.pdf?sequence=1

UNESCO (1982) New trends in Primary Schools Science Education. Paris: UNESCO.

United Nations Educational, Scientific and Cultural Organization (UNESCO) (2009). *Technical paper No.2 Guide to measuring information and communication technologies (ICT) in education.* Montreal: UNESCO Institute for Statistics.

Wellcome. (2022). *From pandemics to prevention: our vision for Infectious Disease* https://wellcome.org/news/pandemics-prevention-our-vision-infectious-disease

Wellcome Trust. (2014). Primary science: is it missing out?: recommendations for reviewing primary science. London: Wellcome Trust. https://wellcome.org/sites/default/files/primary-science-is-it-missing-out-wellcome-sep14.pdf

Wellcome Trust (2017) A Review of Ofsted Inspection Reports: Science. London: The Wellcome Trust https://cms.wellcome.org/sites/default/files/review-of-ofsted-inspection-reports-2017.pdf

Wellcome Trust (2017). *The science content in primary Initial Teacher Training (ITT).* https://wellcome.org/sites/default/files/science-content-in-primary-initial-teacher-training.pdf

Wellcome Trust. (2020) Evaluation of the primary science campaign. (2020) London: The Wellcome Trust. https://cms.wellcome.org/sites/default/files/2020-10/evaluation-of-the-primary-science-campaign-2020.pdf

Wiliam, D. (2011*). Embedding formative assessment*. Bloomington: Solution Tree Press

Wilson, H. Holligan, B., & Hanley, P. (2018) Thinking, Doing, Talking Science. *Primary Science* 154, 15-18. Association for Science Education

World Health Organisation. (2020). Antibiotic resistance https://www.who.int/news-room/fact-sheets/detail/antibiotic-resistance

Appendix 1: Review of literature on practice

A literature review of research on practice in primary science teaching in England

November 2021

Introduction

This non-systematic review of literature focuses on contextual factors that influence teaching and learning science in primary schools in England (age 5-11/Year 1-6) – referred to hereafter as 'primary science.' Intended to map out the field in preparation for the systematic review, it draws from academic literature published since 2013 – that is, literature published since the introduction of the most recent iteration of the National Curriculum for England – and grey literature. The grey literature includes practitioner-focused journals and reports including those by Ofsted. As a narrative literature review that presents the 'state of the art' (Grant & Booth, 2009) of primary science teaching, its chief concern is on current matters; however, some reference to previous curricula or models of assessment is included to offer useful context. Whereas the review of theory examines the theoretical body of knowledge that exists in relation to primary science learning, the aim of this review is to consolidate an understanding of current practice in primary science and to contextualise the outcomes of the systematic review.

A range of factors that influence decisions about the teaching and learning of primary science were identified in the literature. Although the scope of these factors extends beyond policy and decisions made at a teacher or school level, the ways in which individual schools respond to these wider influences can, in part, determine engagement with and enjoyment of primary science and reflect the status that the subject is given in the curriculum. The findings can be organised into the following connected themes:

- Monitoring and assessment: statutory testing (and the absence of this); sampling; formative monitoring, assessment and testing; summative judgements; whole-school approaches and tools.
- Curriculum design and progression: curriculum aims; focus of content; supporting materials; cross-curricular links; progression including transitions in learning from Key Stage 2 to Key Stage 3.
- Teaching approaches: inquiry-based; practical work; discussion; cognition and thinking skills; creativity and drama; self-regulated learning.
- Societal, national and global influences: how the wider community engages with primary science; attitudes and dispositions; the development of science capital; role models and stereotypes; future career aspirations; application and relevance of science.
- Teacher knowledge, development and support: teacher subject knowledge and confidence (Initial Teacher Education, ITE and Continuing Professional Development, CPD); science subject leads; the role of professional organisations and universities in supporting primary science; networks including school clusters and social media.

A wide literature is associated with each contextual theme, however, for the purpose of this review, it is largely only science-specific sources that have been included. Together with the school's ethos and policy approach to teaching and learning (often shaped by the school's Senior Leadership Team), the literature illustrates how the range of influencing factors can shape the practice of science education. What emerges is a fragmented picture of how primary science is organised within and across schools in England with a combination of ideas and resources used to inform practice. Whilst this fragmentation does not necessarily lead to incoherence, the literature suggests that schools do not always have a clear rationale for why certain decisions are made. This highlights the importance of evidence-based approaches to curriculum design, teaching approaches and assessment strategies.

This review will present the literature associated with each theme in turn. The narrative on current practice includes reference to enablers and barriers to teaching science and concludes with the potential of bridging research and practice in primary science.

Theme 1: Monitoring and assessment

The reason for starting this review with a focus on assessment is because of the ways in which the removal of statutory testing has impacted on the status and practice of primary science. In 2009, amidst concerns about an excessive testing culture and too much emphasis on teaching to the test, the then Labour Government abolished Standard Attainment Tests (SATs) for science for 11 year-olds in England. The Association for Science Education (ASE) endorsed the removal of tests and a move towards teacher assessment (ASE, 2016) acknowledging the opportunities this would allow for teachers to be more creative in their science lessons. This seemed like a positive shift to focus on processes and practical work rather than prioritising assessed outcomes.

However, by 2014, there were reports that the removal of statutory tests coupled with greater emphasis on outcomes for English and mathematics had led to a diminished status for science in the primary curriculum (Wellcome Trust, 2014). Perhaps the most significant impact of this has been the reduced time allocated to science learning each week. A relatively recent ‘State of the Nation’ report from the Wellcome Trust (Leonardi *et al.*, 2017) reported primary schools as having an average of one hour 24 minutes of science on the timetable each week with only 30% of respondents thinking that the subject was important to their Senior Leadership Team, compared with 83% for English and 84% for mathematics.

George (2018) asked ‘why is primary science dying?’ and reported that primary science ‘has become a victim of English and maths’ in a criticism that the subject was being squeezed out of the curriculum despite the government’s insistence that STEM skills were key to the UK’s future prosperity. Ten years after the abolition of statutory testing, Parry *et al.* (2019) asked ‘Where has all the science gone?’, questioning whether science is still a core subject on the primary curriculum. Although the latest version of the National Curriculum (NC) for England includes it as such, evidence suggests that this is not translated to practice in terms of curriculum design and the allocated space in the timetable. However, Ofsted’s new inspection framework, which took effect from September 2019, mentioned high-quality education and a connected curriculum which renewed the importance of a breadth of subjects. Furthermore, the ‘broad and balanced curriculum for education recovery’ (Ofsted, November 2021) emphasises the importance of identifying the content in biology, chemistry and physics as ‘most important for enabling pupils to build up their knowledge of key scientific concepts,’ (p.34).

A review of literature concerned with curriculum design and progression will be discussed as the next theme. The remainder of this section will focus on the nature of assessment including teacher assessment, statutory summative judgments and the equity gap in learning exacerbated by the COVID-19 pandemic.

To replace summative outcomes as assessed through the SATs, between 2009 and 2015 summative teacher assessment consisted of ascertaining a level for each pupil in science, continuing the system introduced in 1988 by the Task Group on Assessment and Testing (DES, 1988). This has subsequently been replaced by a system based on age-related expectations with the National Curriculum serving as the target for all children. The following paragraphs present a brief overview of assessment practices in science that emerged since the SAT tests were withdrawn.

In order for the Government's Standard and Testing Agency to monitor overall national performance, a sample of pupils (10 pupils per school from a sample of 1900 schools) sit tests in science on a biennial basis. Despite the ambition for the 'New primary curriculum to bring higher standards in English, maths and science' (DfE, 2012), only 28% of 11 year-olds tested had achieved the expected standard in 2014. Comment on these test results included criticism that children had not been prepared for the tests and that the outcomes did not necessarily indicate children's capacity in science (Robertson, 2017). In 2016, this dropped to less than a quarter and in 2018 reduced further. The 2020 tests did not take place due to the COVID-19 pandemic and the Government stated that there would be no sampling tests in the academic year 2020/21 (gov.uk, 2020).

Alongside the end of Key Stage 2 tests, teachers were required to provide levels of achievement for their pupils linked to NC objectives. These judgments could be informed by continuous 'Assessment for Learning' identified as 'the process of seeking and interpreting evidence for use by learners and their teachers to decide where learners are in their learning, where they need to go and how best to get there,' (Assessment Reform Group, 2002). Generic principles underpinning 'Assessment for Learning' (Black & Wiliam, 1998; Clarke 2001) could be applied to science learning and included the prioritisation of talk, questioning, feedback, self- and peer-assessment together with a focus on identifying and addressing misconceptions (Hodgson, 2010). Although teachers had been involved in assessing learning in primary science, greater significance was placed on their summative judgments when statutory tests were removed.

Davies *et al.* (2014) noted that with no government prescription for assessment of primary science, a diversity of approaches emerged. Whilst they reported the use of a range of creative approaches to formative assessment, they recognised that the lack of centralised policy was having a detrimental effect on summative assessment and the tracking of pupil progress. Earle and Turner (2020) noted that the removal of statutory tests instigated change in school practices and that no support was offered for teachers' assessment literacy concerning how meaningful assessment could take place. This was particularly problematic for teachers who had become entrenched in thinking of learning only in terms of levels. Concerns were raised that use of the criteria set out in the NC could lead to a tick box approach resulting in surface learning rather than devoting time to deeper learning (Black & Wiliam, 1998). There were also concerns regarding the reliability of teacher assessment. Although a range of resources to track and record pupil progress emerged, for example SNAP Science assessment (Beverley *et al.*, 2016) and PLAN assessment matrices (Primary Science Education Consultancy *et al.*, n.d.), questions regarding the relevance and validity of judgements remained.

Initiated in 2011, a now well recognised national programme that aimed to 'champion good practice and raise the profile of primary science' (Turner, 2016) is the Primary Science Quality Mark (PSQM). Supported by the University of Hertfordshire and the Primary Science Teaching Trust, the PSQM is a school improvement programme for primary science, which provides professional development for around 600 new and experienced subject leaders each year, designed to have a whole school impact on science teaching and learning including effective assessment strategies. Earle and Turner (2020) revisited evidence from the PSQM to explore what had happened to teachers' assessment of primary science. Their findings demonstrated how the schools that engaged with the PSQM employed a greater number of Assessment for Learning (AfL) strategies with evidence that some teachers changed their lesson plans to fit new assessment approaches. Time that was previously given to carry out end of topic tests was reduced and greater emphasis was placed on ongoing modes of assessment. Over 4000 schools have achieved a Primary Science Quality Mark (Turner, 2016). PSQM is recognised by Ofsted and other organisations in science education as an effective vehicle for whole school improvement in science.

Another recognised initiative focused more specifically on assessment is the Teacher Assessment in Primary Science (TAPS) project – an approach to assessment that aligns with cognitive theory using repeated, low-stakes assessments and feedback as an effective tool for learning and retention. As an ongoing process, TAPS can be used to identify and address misconceptions. Based at Bath Spa University and funded by the Primary Science Teaching Trust (PSTT), the project aimed to 'develop support for a valid, reliable and manageable system of primary school science assessment which will have a positive impact on children's learning,' (Earle *et al.*, 2017, p.2). It adopts AfL principles and builds on a model for whole school learning first introduced by the Nuffield Foundation (Harlen *et al.* 2012). Earle (2015) noted the importance of adopting whole school approaches to the assessment of science and identified four key features for effective practice: embedding assessment into the planning process; encouraging children to take responsibility for their learning; ensuring assessment is an ongoing process; and offering a clear understanding of what good science looks like across the school. Working with schools nationwide – including those engaged with the PSQM and primary schools in which science is taught by fellows of the PSTT – TAPS provides resources which primary schools use to critically examine the ways in which they use teacher assessment in science. To support decision making in whole school assessment processes, findings from the TAPS project (Davies *et al.*, 2014) were used to create a School Evaluation tool (Earle *et al.*, 2017). The resulting data-flow pyramid is a whole-school self-evaluation tool to identify strengths and weaknesses in primary teacher assessment of science. Used by science leaders and teachers, the tools can be used to plan for greater quality in teacher assessment. The tool is underpinned by a theoretical model of how a whole-school system for the collection, feedback and summary of pupils' science learning assessment data – for both formative and summative purposes – could fulfil the quality criteria of validity, reliability, manageability and impact (Davies *et al.*, 2017). This demonstrates the value of adopting a design-based research approach to school self-evaluation in relation to assessment in science.

By 2018, the TAPS pyramid tool had been downloaded thousands of times across 45 countries (Hopwood-Stephens, 2018). Subsequent analysis of user responses and the ongoing take-up of this tool demonstrated that schools did want to engage with effective assessment of children's learning and, in the absence of government directives, valued resources available from other sources. Emerging issues identified how support for teacher assessment literacy was required alongside any introduction of new assessment systems (Earle & Turner, 2020). The ways in which school leadership and science leads are instrumental in providing such support is presented in the final theme.

A new imperative for schools to engage with meaningful assessment processes is the challenge of learning loss caused by the COVID-19 pandemic and the subsequent school closures for most primary aged children. Canovan and Fallon (2021) noted that barriers during the first lockdown included concerns about whether families could provide resources, poor access to technology and parents' ability to support science learning. A new barrier during the second lockdown was the increased prioritisation of English and mathematics. They reported that schools offered no, or very little, science during lockdowns. They reported that some teachers were told not to include science in the live lessons but to play videos related to science instead. Ofsted (2020) reported that the majority of primary school leaders prioritised reading and mathematics with very few schools focusing on science. Canovan and Fallon (2021) questioned whether learning loss in primary science would simply be forgotten. They raised concerns associated with inequality stating that 'science teaching had suffered in ways that were likely to entrench inequality in who can access STEM education and careers' (p.46). Their study reported that teachers were worried that reversing science learning loss was not a priority for some schools, and/or reported that no efforts had been made to tackle this.

This loss of learning is likely to impact on pupils of low socio-economic status who have already been identified as disadvantaged in science learning (Nunes *et al.*, 2017). A priority resulting from the identification of the extent of lost learning thus includes the need for teachers to carefully assess the learning gap and to consider how future learning might incorporate what was missed. Redhead (2021) suggested a developing approach to mapping the curriculum post-COVID-19 pandemic and included an emphasis on formative assessment strategies to establish children's understanding.

Removal of statutory tests in 2009 provided an opportunity for primary teachers to focus on science learning rather than test preparation. Despite science being named as a core subject in the NC, and despite the ongoing necessity for teacher assessment of learning, the status and lesson time for primary science has largely diminished. Given the limited time planned for science each week, the emerging question focuses on what content and skills should be learned and how this is best achieved. The next section presents aims of various curriculum materials used in schools, how cross-curricular links might enhance science learning and the challenges associated with transition from Key Stage 2 to Key Stage 3.

Theme 2: Curriculum design and progression

Analysis by the TES (Bloom, 2017) raised concern about the narrowing of the curriculum highlighted by the lack of focus on science in the key findings section of Ofsted reports. References to science appeared in only 4% of reports. Bloom noted that it is difficult to discern what importance the Government wanted to place on science and raised questions about the extent to which Ofsted inspectors considered a school's offer of a broad and balanced curriculum. Wellcome (2015) reported that science was mentioned in only 27% of Ofsted's primary school inspection reports during the previous year. Documented in later reports by Wellcome, this rose to 61% in 2017/18 (Wellcome, 2018). By 2019, Ofsted had planned a research review of science which (due to delays caused by the Covid-19 pandemic) was published in 2021.

In many primary schools, teaching the content of science objectives in the previous national curriculum (Department for Education and Employment, 1999) was largely shaped by schemes of work for each topic. These were provided by the Qualifications and Curriculum Agency (established in 1997 and dissolved in 2011) who had a role to monitor and review the core national curriculum subjects to ensure that the quality of curricula remained high. The current National Curriculum (DfE, 2013) was designed as a more 'knowledge-rich' curriculum which placed a greater emphasis on science knowledge and included some topics which had previously been taught at Key Stage 3.

The National Curriculum for science (DfE, 2013) aims to ensure that all pupils:

- develop scientific knowledge and conceptual understanding through the specific disciplines of biology, chemistry and physics
- develop understanding of the nature, processes and methods of science through different types of science enquiries that help them to answer scientific questions about the world around them
- are equipped with the scientific knowledge required to understand the uses and implications of science, today and for the future.

Parry *et al.* (2019) suggested that there should be a reduction in the content of the NC to provide more time to focus on 'working scientifically'. They argued that this would 'allow topics to be explored in a creative and exciting manner with time being allowed for whole investigations and deeper exploration of key science ideas'. They suggested that this would 'lead to greater motivation, enthusiasm and love for the subject and the development of key thinking skills,' (p.14).

The current curriculum is not supported by any teaching materials from the DfE, leaving schools with the autonomy to decide how to meet curriculum objectives – and for schools with academy status which are not required to follow the national curriculum, greater autonomy still. Barnes (2015) noted this political freedom and argued that it provided an opportunity to rethink the curriculum advocating the connection of learning through cross-curricular approaches. However, with this freedom came the question of how schools were equipped to organise their curriculum without compromising the quality of learning. Roberts (2019) reported that 'when mapping the national curriculum against some of these published curriculum resources it is alarming that the key content is not always covered and opportunities for promoting working like a scientist are often lacking'. With the body of knowledge compiled by the QCA gone, Earle and Turner (2020, p.5) recognised that some teachers may have been left 'lacking supportive statutory structures and perhaps an uncertainty regarding why and how to teach primary science.'

Since 2013, a number of curriculum schemes and initiatives have been introduced, written by – or in partnership with – publishers, subject associations, private companies, funding bodies and universities. The purpose of this review is not to assess the available materials. However, noting the resources' purpose and aims highlights what factors might shape the decisions made by schools when planning their science curriculum. That is, how does a school organise their curriculum to meet the above aims and teach the content listed in the national curriculum? The wider role that various organisations play in primary science extends beyond curriculum materials and will be referred to in the theme focused on teacher knowledge and CPD. Some of the materials cover the whole curriculum and others focus on ideas for practical work. Resources include: Explorify, PSTT, Wow Science, Thinking Doing Talking Science, BBC Terrific Scientific, Empiribox Primary School Science, Oak National Academy and Best Evidence Science Teaching (BEST). Appendix 1A presents more detailed information about each resource including the aims and the relevant weblinks.

It is clear to see that primary schools are well-resourced with teaching materials, however, without a clear rationale underpinning curriculum design there is danger that science could become a set of disconnected activities. This highlights the need for a carefully planned curriculum design that incorporates progression and assessment of learning. Based on their mission statement or philosophy, different schools, organisations and educators will emphasise different factors important in curriculum design. The literature references a number of overarching purposes for teaching primary science that can influence curriculum design.

The Teaching Schools Council's report (Keeble, 2016) identified that in the most effective schools, 'the vision for teaching helps pupils foster a love of learning that is connected to the wider context' (p.11) which it claimed can develop pupil vocabulary including the technical vocabulary required to understand science. Wilkinson and Stallard (2019) discussed 'mastery' in primary science which included a focus on the development of conceptual understanding, teacher questioning and rich learning tasks. They argued that enabling children to apply, analyse, evaluate, and/or create to solve exciting and novel problems supports the development of mastery level knowledge and skills in primary science. The Royal Society (2021) recognised the importance of getting science education right and set out a vision for a broad and balanced curriculum up to the age of 18, including mathematics and science. They noted that 'inspirational mathematics and science lessons should be at the heart of the curriculum, and there should be an emphasis on practical work and problem solving'. Similarly, Wellcome (2021) stated that primary science should nurture children's natural curiosity, allow them to ask questions and to develop the skills they need to answer these questions. They highlighted how primary science helps children to investigate problems, learn how science works and discover why science matters in the world. The PSTT (2021) has argued that because attitudes towards science are shaped in primary school, there is a need to ensure that children do not lose interest and enthusiasm for the world around them and the role that science plays.

These points raise questions about how scientific skills and processes integrate with content knowledge and the nature of knowledge. In asking 'what is science?' Crompton (2013) cited the work of Eshach and Fried (2005) who distinguished between conceptual knowledge as developed through science activities that help children to interpret and make sense of the world in which they live, and procedural knowledge which includes reasoning skills such as asking questions, hypothesising and analysing. How these forms of knowledge are taught and assessed varies across schools. Analysis of school data (Earle, 2014) identified that teachers separated scientific skills and knowledge, particularly in relation to summative assessment, which called into question how the nature of science was represented in schools and whether it was 'possible or desirable to separate knowledge and skills in this way' (p.225). Another differentiation of knowledge outlined in Ofsted's (2021) research review identified: substantive knowledge (knowledge of the products of science, such as concepts, laws, theories and models) referred to as scientific knowledge and conceptual understanding in the national curriculum; and disciplinary knowledge (knowledge of how scientific knowledge is generated and grows). This is included in the 'working scientifically' sections of the NC and includes knowing how to carry out practical procedures.

In the programme of study for the national curriculum (2013), 'working scientifically' is outlined separately from but 'must always be taught through and clearly related to the teaching of substantive science content in the programme of study', emphasising the way in which children learn scientific concepts. Smith (2016) noted the change in terminology of 'scientific inquiry' (pre 2000 curriculum), 'how science works' (post 2000 curriculum) and 'working scientifically' (post 2013 curriculum). Smith (p.4) highlighted the importance of identifying the various skills that comprise 'working scientifically' citing Harlen and Qualter (2015), who classified four areas: raising questions, predicting and planning; gathering evidence by observing and using information sources; analysing, interpreting and explaining evidence; and communicating, arguing, reflecting and evaluating evidence. Mepsted (2018) suggested that when first introducing these skills, 'a single skill should be focused on in a single science lesson' (p. 19) so that children can understand the components that make up each skill and to support effective assessment of children's skills.

Harlen *et al.* (2010) outlined 'principles and big ideas of science education.' Harlen *et al.* (2015) argued that these ideas of and about science (identified by international science education experts) underpin the curriculum and 'have explanatory power in relation to a large number of objects, events and phenomena…; provide a basis for understanding issues…involved in making decisions that affect learners' own and others' health and wellbeing and the environment; lead to enjoyment and satisfaction in being able to answer or find answers to the kinds of questions that people ask about themselves and the natural world; and have cultural significance' (p.99). The implications that the big ideas have for curriculum content highlight the importance of 'ensuring that students' learning in science gradually builds into a coherent whole, and is not left as a set of disconnected facts,' (Harlen *et al.*, 2015 p.47).

The literature identified two ways to promote more time for primary science by organising the curriculum differently – through cross-curricular links and whole school science weeks. Although sometimes prevented by setting or streaming across classes, the nature of the organisation of teaching and learning in a primary school can lead to greater flexibility with the timetable thus allowing the teacher to make explicit links between subjects, knowledge, understanding and language. Wider debates in curriculum theory regarding bounded knowledge domains versus connected learning approaches is beyond the remit of this review, however, the literature identified how meaningful connections between subjects can contextualise learning and provide opportunities to apply understanding and skills in other curriculum areas, for example through STEM (Science, Technology, Engineering and Mathematics) projects and connecting with the Arts (STEAM). One of the goals of STEAM approaches to curriculum design has been defined 'to increase the participation of students who are traditionally absent from STEM' (Quigley & Herro, 2016 p. 422). This illustrated how children who may have already dismissed science as a subject they are interested in could be re-engaged through cross-curricular design.

CLEAPSS (set up as an advisory service to support science and technology in schools now recognised as a name and no longer as an acronym) was founded to promote high-quality, effective practical work in science. Their website is designed to support practical science, D&T and art including information for primary teachers and leaders of science, technology, art and design. It includes doing things safely, primary competitions and leadership guidance. The bank of teaching ideas links learning across subjects.

Borthwick and Cross (2015) recognised the benefit of integrating the skills of working scientifically and working mathematically even though both are valuable in their own right. Markwick and Clark (2016) identified the ways in which children's understanding of scientific concepts and ideas could be achieved by exploiting links between science and mathematics. Planning for synergy between these subjects also provided a range of contextualised opportunities. Markwick (2019) demonstrated how children's experience of learning science could be enriched when application of mathematics and English were integrated into lessons. Although his focus considered how mathematics mastery could be assessed by working scientifically, this approach to planning had potential for dedicating more time to science in the curriculum. One approach to this is described by The Paradigm Trust. In their schools, one week each half term is devoted to what they describe as 'disciplinary literacy' lessons (Rogers, 2021) which replace English lessons. These focus on scientific English in which pupils read, discuss and write about the science they have been learning. In addition to increasing the time for science on the timetable, these approaches to curriculum design demonstrate that children's experience of learning science can be enriched when connected to other subjects. Whole school science weeks also promote cross-curricular links and provide extended time for science learning. Bostrom (2016) outlined how extended time enabled child-led inquiry where children experienced different steps in a scientific investigation.

Another influence on the curriculum design for primary science concerns the transition to secondary school. Ofsted's report 'Key stage 3: the wasted years?' (2015) which investigated whether KS3 was providing pupils with sufficient breadth and challenge had very little detail about science education. Data on the extent to which pupils said that they were doing the same work in Year 7 science lessons as they had done at primary school was presented with foundation subjects as a combined figure ranging from 9-17%. This data was presented separately for English and for mathematics. Ben Rogers (director of curriculum and pedagogy at the Paradigm Trust MAT) noted the finding by Wellcome that around half of students in Year 7 and 8 felt unprepared for science at secondary school (Wellcome, 2020, p.7). Having experience of teaching as a secondary school science teacher and more recently as a primary teacher, Rogers commented that scientific knowledge was not sufficiently prioritised at primary school claiming that 'if you look at published resources and training for primary science, the emphasis is on enquiry teaching,' (Rogers, 2021). With what he referred to as a more instructional approach at Key Stage 3, the MAT adapted primary science education to lay a preparatory foundation for science learning at Key Stage 3. Barnes (2015) referred to how some imaginative primary and secondary schools had collaborated to form a 'bridging curriculum' sharing themes – which are often around science – and introduced children to the resources and facilities for science at secondary schools.

Whilst there are extensive resources and activity ideas available to primary teachers (Appendix 1A), the literature demonstrates the importance of engaging with curriculum purpose, design and progression. Leach (2021, p.69) noted that 'it is critical that science teachers have professional autonomy and agency to contribute to discussions on the purposes of science education and the design and implementation of approaches to realise those purposes through the curriculum'.

Theme 3: Teaching approaches

Whereas the previous theme focused on issues of curriculum design and overall aims, this theme presents literature focused on different approaches to teaching and learning in primary science. This includes child-led inquiry-based approaches, practical work, problem-solving, the role of talk, discussion and questioning, development of thinking skills, self-regulated learning, and creative approaches including drama and getting into a role. These examples of practice are not linked to an underpinning theoretical framework, as this is explored in depth in the theoretical review, but they do highlight how decisions made about pedagogical approaches can influence children's learning of science.

Based on the concept of active and experiential learning, inquiry-based approaches are commonly included in primary science materials. The PSTT identifies a number of different types including: comparative/fair testing; research; observation over time; pattern seeking; identifying, grouping and classifying; and problem solving, often involving practical work.

Harlen (2013) noted that inquiry can be applied to a number of curriculum subjects but that 'what distinguishes scientific inquiry is that it leads to knowledge and understanding of the natural and made world through direct interaction with the world and through the generation and collection of data for use as evidence in supporting explanations of phenomena and events' (p.12). This understanding of the nature of inquiry highlights the importance of planning for learning rather than just planning for practical activities. Harlen (p.16) emphasised the 'process of building understanding through collecting evidence to test possible explanations and the ideas behind them in a scientific manner' thus integrating scientific skills with conceptual understanding. This 'hands-on' and 'minds-on' approach to practical work was promoted by Abrahams *et al.* (2014) who emphasised the importance of planning which links observations and scientific ideas, and of open-ended investigations rather than finding pre-set outcomes through 'recipe' style approaches to practical work.

McCrory (2017) outlined how inquiry-based approaches could motivate and engage learners and argued that exploration and inquiry are 'crucial in developing process skills for children to construct their understanding of conceptual science,' (p.8). Whilst harnessing curiosity and fostering a love of science through fun activities, McCrory noted that such approaches must be designed to promote progression of conceptual understanding, scientific thinking and reasoning skills rather than simply being a fun activity. This highlights how unfocused teaching pedagogies can potentially become barriers to learning when explicit links between scientific concepts and the inquiry undertaken are not made.

Through the 'Thinking, Doing, Talking Science' (TDTS) project, Wilson *et al.* (2018) emphasised that primary science activities can be simple but need to be effective. This could be accomplished through a design for practical science that encourages children's learning and engagement through creative, interactive and cognitively challenging lessons. Development and employment of specific higher order thinking skills (based on concepts such as Bloom's Taxonomy, 1956) and dedicated time for discussion are component elements of this approach. Hanley *et al.* (2020) reported on the impact of CPD in TDTS to develop teacher knowledge, confidence and ability to increase the level of conceptual challenge in primary science. Questioning through 'doing' (minds-on practicals and focused recording) and talking (specific tasks and activities) supported and enabled the development of children's higher order thinking skills. Linking carefully selected materials with clear learning opportunities is also a feature of Concept Cartoons. Spanning a number of years, research by Naylor and Keogh (2015) identified their value in the promotion of cognitive conflict and argumentation, as tools for formative assessment, challenging misconceptions and enhancing motivation and engagement.

An approach designed to inspire deep learning and to capture children's imagination used inventive activities and stimulating contexts. Cutting and Kelly (2014) recognised the value of using creative approaches in primary science and McGregor (2014) noted how a new creative space was opened up for primary science when statutory tests were removed. She found that working with drama in learning increased children's engagement, involvement and motivation as well as supporting their conceptual and procedural understanding.

Using drama-based approaches with children acting as scientists-in-role, McGregor *et al.* (2019) identified how children could better appreciate the nature of science by providing opportunities to be agentive and to think and act scientifically. By drawing on principles underpinning a curriculum approach 'Mantle of the Expert' (MoE), they adopted a Learning Science through Drama (LStD) approach to create 'as-if' worlds. The MoE approach (Heathcote & Bolton, 1995) is rooted in three pedagogic structures: inquiry learning; drama for learning; and expert framing. Through an emphasis on self-directed research and discovery, children are empowered as expert learners. The findings by McGregor *et al.* (2019) outlined how the application of drama conventions could offer a pedagogy that had the potential to support learning about the dispositions required of scientists, the nature of science and the development of scientific thinking and practice.

Maxwell *et al.* (2019) also combined elements of the MoE dramatic-inquiry approach with design thinking and design fiction to explore environmental issues. With the intention of stimulating creativity, their approach established a fictional but believable world in which children could engage with environmental challenges. They reported that through a learner-centred, design-approach to the project, children demonstrated high levels of engagement and enthusiasm and were able to imaginatively think through solutions to complex issues associated with global environmental threats. Stagg (2020) meanwhile espoused the use of drama in primary science to increase motivation, enjoyment and interest. Through participation in inquiry-based learning, drama games and authentic problem-solving activities in primary science, Stagg demonstrated measurable positive gains in both learning and attitudes.

This range of teaching approaches demonstrates ways in which teachers seek to engage, excite and motivate learners through a variety of activities to stimulate thinking and discussion. Emerging implications identify how 'hands-on' learning also stimulates 'minds-on' engagement and the importance of making explicit links between the activity and the intended learning.

Theme 4: Societal, national and global influences

The literature demonstrates how local, national and global initiatives and resources can influence the teaching and learning of science in primary schools. Some of the reported drivers for a wider engagement with scientific ideas are children's disinterest in science, a lack of aspiration for future careers in STEM jobs (Archer *et al.*, 2020) and the misinformed stereotypes of what being a scientist means (Davenport & Shimwell, 2019). In the first ASPIRES project, Archer *et al.* (2013a) found that attitudes to science were largely fixed by the end of primary school and that students' aspirations were shaped by the amount of 'science capital' that their family had. They identified this as 'science-related qualifications, understanding, knowledge (about science and 'how it works'), interest and social contacts (e.g. knowing someone who works in a science-related job),' (p.3).

In their article 'Not girly, not sexy, not glamorous – primary school girls' and parents' constructions of science aspirations' (Archer *et al.*, 2013b) noted that the girls were less likely to choose science related careers due to constructions that such careers were 'clever/brainy', 'not nurturing' and 'geeky' (p.171) which did not align with girls' self-identifications. Oliver (2013) noted the importance of humanising science and humanising scientists in primary science and that 'by looking at the lives of scientists pupils will be helped to recognise science as a human endeavour, not the activity of caricatured stereotypes' (p.32). In its review of what pupils (across the UK) think of science in primary schools, the Wellcome Trust (Leonardi *et al.*, 2019) noted that 29% of pupils reported a science related job when asked what they would like to do in future. Perhaps this was not surprising when the review also reported on attitudes towards science. When asked how much primary school children like science at school, 44% liked it a lot, 42% liked it and 14% didn't like it.

A report published by the Royal Society and CBI (2016) highlighted how knowledge about careers and the potential opportunities and pathways available to young people was important while they were building their identities. They noted research from King's College London which showed that if 10 year-olds could not visualise themselves as a future scientist or an engineer, then they were unlikely to be able to do so by the time they were 14. They therefore emphasised the importance of positive STEM business interactions and a diverse range of positive role models in primary science. Perhaps in an attempt to help children to visualise themselves as scientists, Ridley (2014) noted how children in her primary school wore white lab coats whenever a science lesson was taking place. Intended to raise the profile of science across the school, this was not balanced with caution about perpetuating stereotypes. Davenport and Shimwell (2019) advocated the need for careers education in primary schools to counter the impact of gender and socio-economic status on children's career ideas.

Connections with science outside of the primary school were of particular importance for those children who had low science capital, defined by Chowdhuri *et al.* (2021) as what a child knows about science, what they think about science, what science-related activities they engage with in their spare time, and who they know who might promote engagement with science talk or activity. In a project led by researchers at UCL and King's College London, Chowdhuri *et al.* (2021) argued that there is an imperative for primary science to be socially just so that all children could be empowered to make informed decisions about their wellbeing, environmental issues and their future career. Partnering with teachers to integrate equitable engagement of children with science, they launched the Primary Science Capital Teaching Approach (PSCTA). This approach, designed to sit alongside a school's curriculum, focused on inclusive teaching and learning and supporting children's voice and agency. The impact this approach made on primary science included: increased science identity; science trajectory; out of school engagement and science agency.

The merits of engaging children with science beyond school-based learning opportunities include challenging attitudes about the role and value of science, empowering children to ask questions and follow their own lines of inquiry, providing opportunities for first-hand experiences and seeing the relevance of scientific concepts through application to real-world settings. That is, such experiences can help to develop a child's science capital. The reviewed literature demonstrates how teachers link opportunities in primary science with a range of initiatives and resources that shape children's curiosity, enjoyment, attitudes and aspirations.

An example of government involvement in developing primary science beyond school-based practice was through targeting specific communities. Stoke-on-Trent Opportunity Area was launched to raise education standards, providing every child and young person in the area with the chance to reach their full potential in life. Included in this was the initiative 'Science Across The City' (SATC), designed to raise the profile of primary science, support strategic development of primary science, engage professional learning communities in key current issues in primary science and to connect schools to additional resources and wider opportunities through the STEM landscape. In partnership with the PSQM and PSTT, the initiative sought to narrow the gap in opportunities between different schools in order that every child experienced child-led, inquiry-led science for success. The Evaluators' Report (Warren, 2021, p.22) identified that compared with other science subject leaders undertaking the PSQM, the extent of engagement with primary science networks was 'unprecedented' within the SATC communities.

The House of Commons Science and Technology Select Committee (2017) recognised how initiatives that complement science learning in formal education could increase science capital through encouraging young people to engage with science. One such example (Hobbs *et al.*, 2019) demonstrated how the computer game Minecraft served as a hook to engage primary aged children with science concepts through a project 'Science Hunters' launched by the University of Lancaster in 2014. This outreach programme used engaging projects to communicate scientific concepts and inspire children's interest in, and enthusiasm for, science.

National initiatives such as British Science Week (British Science Association, 2023) and associated 'citizen science' projects can link well to primary science and are supported by tailored downloadable resources. Designed to promote engagement and link to the curriculum, Tyler (2019) noted that Citizen Science projects could demonstrate the relevance of science to children's lives helping them to understand how and why science research is being carried out. He noted how this approach could offer everybody the opportunity to get involved in ongoing real scientific studies by collecting or analysing data and uploading their results. Partnering with commercial and charitable organisations, projects have included digging up squares of soil to count earthworms, examining lichen on trees and digitalising handwritten meteorological data from 1860-1880. Representing a number of universities around the world, Roche *et al.* (2020) recognised that collaboration with citizens in scientific research was a growing field that generated new knowledge and understanding and had the potential to foster education and learning opportunities.

Growing from children-to-children's science conferences, the Great Science Share for Schools (GSSfS) was launched in 2016. This campaign was designed to engage with school communities to raise the profile and engagement of young people in primary and secondary school science. Engaging 'everyone who cares about young people and their understanding of the world they live in,' the campaign was underpinned by three foundational values: child-focused science communication; inclusion and non-competition; and promotes collaboration. With links and connections to the wider community and internationally, the profile of science was raised and supported the development of science capital. Bianchi (2021) reported that the campaign resulted in more time for science learning in school and at home giving opportunity for young people to ask, investigate and communicate their own scientific questions with new audiences on a national platform.

Trew *et al.* (2023) identified how research could be introduced into primary science. With varying levels of adaptation, they demonstrated how research articles (in this study those specifically associated with biological sciences) could be meaningfully disseminated to primary aged children and how ‘children can carry out investigations associated with cutting-edge research in the classroom’ (p.1). This illustrated how children could learn to see themselves as scientists connected to real life science (Cross & Board, 2014). Trew *et al.* (2023) highlighted the potential benefit that research engagement could have to children, primary teachers and to the researchers. Having used research articles in their lessons, teachers noted an improvement in children’s subject knowledge and their inquiry skills and that children saw how science could be used in real life situations. They also suggested how science capital might be raised through investigations because children could identify themselves as scientists.

A range of global scientific endeavours such as spaceflight can offer meaningful links with primary science. Tim Peake’s Principia Mission to the international Space Station enabled direct engagement with children suggesting experiments for him to carry out in space. STEM learning (STEM, 2021) provided a range of ‘Tim Peake Project Activities’ supported by a range of funders to promote children’s interest in science. Lister (2018) recognised the enthusiasm that children have for learning about space and demonstrated how primary science could be contextualised through a range of resources and experiences in school including resources from the NASA website, activities using telescopes and stargazing evenings. Dunlop *et al.* (2020) outlined how attitudes towards science could be influenced by the ways in which formal and non-formal providers can work together. Using the focus of the 2015 Principia space mission and the connected education and outreach programme, they explored what factors shaped young people’s attitudes to space. The study drew on actor-network theory which maps relationships between ‘actors’ that make a difference giving equal weight to human and non-human influences. Although focused on one specific aspect of science, the findings demonstrated how enacted policy could positively or negatively influence young people’s attitudes through actors including curriculum policy, film, social media, teachers, friends and family.

Another global theme that has been linked to primary science is environmental issues and climate change. The Eco-schools Green Flag Award, The Woodland Trust’s Eco Green Tree School Award and the WWF’s Sustainable Schools Award are ways in which primary children can engage with global issues on a local level. As part of COP26, the Education Secretary set out a vision to ‘empower all young people in the importance of conserving and protecting our planet, as well as developing the skills needed to solve the problems’ by placing climate change at the heart of education. The Department for Education set out to support teachers to deliver a ‘world-leading climate change education through a model primary science curriculum...in place by 2023 and to teach children about nature and their impact on the world around them’ (DfE, 2022).

Involvement from external bodies such as guest speakers or theatre companies such as GrowTheatre (Maxwell *et al*., 2019) coming into schools, as well as school visits to outside spaces or museums, can promote relevance and context for scientific ideas. With access to technology, many museums and organisations have extended their reach to support learning at school and at home, for example, London’s Institute of Imagination at Home ran creative workshops and programmes to foster creativity and imagination through activities and experiments.

Tyler (2018) recognised that making real life links to other areas of life was important in demonstrating the relevance of science in the real world. The examples presented identified the ways in which local, national and global issues and initiatives could be linked to primary science. The challenge is how to harness the learning potential in meaningful ways that support a well-designed curriculum.

Theme 5: Teacher knowledge, development and support

The final theme focuses on teacher knowledge, teacher confidence and teacher support as contextual factors that influence primary science. This includes Initial Teacher Education (ITE), the role of school science leads, Continuing Professional Development (CPD) professional bodies, universities, charities, private organisations, teacher networks and social media.

Wellcome's report on primary science education (Leonardi *et al.*, 2017) noted that 25% of teachers were concerned that they might not be able to answer children's questions about science. As generalists teaching the breadth of the curriculum, primary teachers have time to get to know individual children well – which can contribute to development of science capital – and have flexibility to connect learning between subjects, however, in 2010, just 3% of primary teachers had a science degree (Royal Society & CBI, 2016, p.13). The importance of good subject knowledge was identified by Hashweh (1985) where teachers with more subject knowledge 'were more likely to detect children's misconceptions and were more likely to 'digress' into other discipline-related avenues, deal effectively with class difficulties and correctly interpret pupils' insightful comments' (p.305). Although now dated research, Hashweh identified a still live issue which identifies specific reasons why teacher subject knowledge is important. It also raises the question whether, over 30 years on, more teachers have a more secure subject knowledge. From their research of the impact of a training programme for primary school teachers, Bennett *et al.* (2019) concluded that whilst knowledge enhancement courses did increase teachers' confidence, they did not necessarily increase subject knowledge.

Given the increase in school-based ITE and a greater proportion of time spent in schools on university-centred ITE programmes, there is reduced time to develop teachers' subject knowledge. George (2018) reported that ITE students questioned the relevance of the content they learned and recognised primary science as feeling like a 'dying field'. Those who did report feeling confident and enthused about teaching science found it difficult once in school given the lack of time and resources devoted to science. Suggestions that university ITE departments could support CPD in a group of schools could strengthen both primary science and inform ITE provision (Parry *et al.*, 2019). McCullagh and Doherty (2018) found that principles of good science education introduced in university-based sessions using a video analysis tool in microteaching activities increased the confidence of pre-service teachers. However, they noted that pre-service teachers' confidence and competence might be limited by the lack of opportunity to observe and teach high-quality science lessons during their teaching placement due to the focus on English and mathematics. With 'very limited opportunities to observe and teach science during a school placement' (p.529), Blackmore *et al.* (2018) identified how pre-service teachers' perceptions of their developing professional identity were negatively influenced. Although Ofsted's education inspection framework (DfE, 2019a) promoted a broad and balanced curriculum which created the potential for schools to move away from the bias towards these two core subjects, Leach (2021) raised concerns about reductions in subject pedagogy in school-led ITE and noted 'the recent Early Career Framework for beginning teachers (DfE, 2019b) adopts an almost entirely generic and technicist approach to teacher development in the early career", noting also that the Early Career Framework is not well aligned with Ofsted's review of evidence on the effective teaching and learning of science. His concern was that beginner teachers would not be equipped to ‘engage critically and influence discussions about the purposes of education and their implementation in the curriculum’ (p.71).

Ofsted’s inspection framework (DfE, 2019a) placed greater expectations on teachers, curriculum and subject leaders to identify the programme of study that year groups were following, the intended endpoints towards which those pupils were working, and their view of how those pupils were progressing through the curriculum. Inspectors would ‘engage in discussion with subject specialists and leaders about the content and pedagogical content knowledge of teachers, and what is done to support them.’ In 2017, it was reported that 52% of science leaders had external CPD for a day or more in the previous year to help them lead or develop science in their school and 30% no support for improving science in their school (Leonardi *et al.* 2017).

Mackintosh *et al.* (2017) found that in most primary schools science was taught and led by class teachers, many of whom expressed a lack of confidence about teaching and leading science. They explored the impact of the PSQM and found that many science leaders reported that they felt better equipped for their own teaching and that the programme helped them to understand their leadership role which made them feel better equipped for leading others. In turn, they reported how the attitude of other teachers changed – both by being inspired to teach science in a more engaging way and to work collectively to develop science across the school. The impact of the intentional development of science leaders was recorded by Wellcome (Leonardi *et al.*, 2019). They found that 48% of pupils ‘like science a lot’ in schools where a science leader had undertaken CPD in the last year compared to 37% where they had not undertaken CPD. 46% of pupils ‘like science a lot’ in schools where respondents to the science leadership survey stated there is a good range of science equipment at their school, compared to 40% where they stated ‘to some extent’.

A rich example of the potential impact that a non-specialist science leader could have on the whole school vision for science was outlined by Orchard *et al.* (2019). Key points included the ways in which science was promoted as a priority, how staff were enthused and excited about science and how teacher confidence was developed. Headteacher support was identified as important including placing science in the school improvement plan, release time for CPD, conferences and the completion of PSQM award. The subject leader used the Forest School and school grounds to intentionally develop science capital for those children who did not have a garden. The primary leader also: started a lunchtime science club where children took part in a CREST award (a scheme managed by the British Science Association that inspires young people to think and behave like scientists and engineers with programmes of challenges tailored to Key Stage 1 and 2); devoted time to researching grants and bursaries to help to develop science in the school; set up class-based science tables; and partnered with STEM Ambassadors to visit the school to talk to the children.

Since the introduction of the new national curriculum, a greater emphasis on teacher assessment, summative statutory sampling and updated Ofsted inspection frameworks, there has been a growth in CPD offered to support primary science. Perhaps in place of the historical role of the Local Authority, there has been an increased engagement of universities and professional subject bodies to develop primary science. Coupled with this has been the development of networks – both face to face through the creation of school clusters and via technological tools including website resources, social media and, more recently, video conferencing.

Insufficient time for subject knowledge and pedagogy on ITE courses, new curriculum content (for example Key Stage 3 topics introduced to national curriculum primary science, 2013) and changes to assessment approaches all lead to a demand for ongoing CPD. A number of universities, professional organisations, funding bodies and other organisations were referenced in the literature and show examples of CPD and support available for primary teachers and subject leaders. These include: the Association of Science Education (ASE), PSTT, PSQM, Primary Science Education Consultancy, The Ogden Trust, the Primary Science Capital Project, Reach Out CPD, the Teacher Assessment in Primary Science (TAPS) project, STEM Learning, Thinking, Doing, Talking Science (TDTS), Children Challenging Industry (CCI, at the Centre for Industry Education Collaboration). Details of the ways in which these initiatives support primary science are included in Appendix 1B together with relevant weblinks. An emerging question that influences the practice of primary science is how schools make decisions about which organisations to connect with, which teaching materials and assessment tools to use and how to discern claims of 'research-based evidence' and its relevance to their setting. The role of the science leader has become instrumental in these decisions.

Specialist science teachers are rare in primary schools (Royal Society & CBI, 2016). The challenge of taking on the role of science leadership may be because teachers have not developed a clear or confident identity as science teachers. Danielsson and Warwick (2014) asked if there was anything specific about being a teacher of science and noted the importance of developing their own scientific literacy. As the only science leader in a primary school, connecting with others from different schools appeared to develop this shared role identity. In addition to CPD programmes, group membership was facilitated through networks and informal communities of practice formed using social media to connect with others in order to share knowledge, ideas and resources and keep up to date with government initiatives that impact on their role.

The appetite for science leaders to connect with others is reflected in examples including ASE TeachMeets (informal, fun and inspiring way for teachers to share ideas) and Facebook communities such as the group for PSTT with over five thousand members. Twitter also connects those working in primary science.

The purpose of briefly outlining the work of universities, professional organisations and private companies highlights an increasing interest in primary science by a number of stakeholders. There is recognition that enjoyment of science, attitudes towards the subject and aspirations of future careers are shaped during the primary phase of formal education. The work of universities and other organisations can therefore play a significant role in engaging teachers with the findings of evidence-based research to shape curriculum design and approaches to assessment, provide resources that have a clear theoretical underpinning and support the ways in which they put this into practice. The final section briefly presents recent research and highlights the potential impact of research on the practice of primary science.

Bridging research and practice – emerging issues

Although reference to research-based approaches to curriculum design, pedagogies, assessment tools and CPD have been included in the five themes identified from the literature, this section focuses on recent research and the associated emerging issues. This includes reference to barriers to children's learning as well as indication of positive trends for primary science practice. Mapping findings from Wellcome reports, Stubberfield (2021) identified how each key indicator for primary science delivery and leadership has improved. Indicators included the average science teaching time, the proportion of schools with a science subject leader and the percentage of science leaders accessing professional development for science leadership or school development. Emphasising the importance of science subject leaders, data indicated that over half of the teachers surveyed recognised that their science leader had provided training, coaching or mentoring to equip them to teach science better. For about a fifth of teachers, lack of teacher confidence was associated with certain topics such as forces and light and over an eighth of teachers lacked confidence in the assessment of science. Stubberfield (2021) concluded that primary science provision was improving slowly. To continue this trend, teachers' understanding about progression of children's learning, the integration of meaningful assessment and increased confidence in subject knowledge is required. The role of the science leader is instrumental in this.

The uptake of evidence-based CPD programmes and assessment tools (e.g. PSQM, TAPS, TDTS) has contributed to improved confidence and leadership of primary science in schools. Their report 'Harnessing educational research' (Royal Society & British Academy, 2018), identified that practitioners' research priorities included 'improving the teaching and learning of science' (p.33). The related web page highlighted how progress had been made towards this since 2014 stating: 'the National Foundation for Education Research and the Education Endowment Foundation continue to lead on the use of research and evidence by teachers; the number of Research Schools expands to 23, aimed at increasing the use of evidence in teaching; the Chartered College of Teaching signs an evidence Magna Carta with other professions in support of evidence-led teaching'. Their website set out their vision to bridge research evidence to shape practice: 'Education policy and practice are better informed by evidence. Education in the UK will benefit from a strong foundation on evidence, and the principle for basing education policy on research needs to be re-established'. They called for greater collaboration between science and mathematics education researchers, scientists, teaching professionals, policy makers and the public.

Using evidence from SEERIH's 'Deep Dives' (Bianchi, 2015) and other tools used by primary schools to review science, Bianchi *et al.* (2021) presented '10 key issues with children's learning in primary science in England'. The significance of identifying these issues lay in the way that children's experience of learning science at primary school 'underpins their identity, ability and the subsequent choices and options that they have for further study in STEM subjects' (p.3). The audience of the report was wide-reaching for all stakeholders in primary science including teachers, school leaders, governors, the STEM community, the inspectorate, policy makers (local, regional and national) and CPD providers. By highlighting these issues, the report identified their impact on learning in primary schools and classrooms and importantly, to 'shed light on where further professional learning and strategic development is required.' (p.12). Given these are recently published findings with relevance for ongoing research, the ten issues are listed here: children's science learning is superficial and lacks depth; children's preconceptions aren't adequately valued; children's science learning lacks challenge; children are over reliant on teacher talk and direction, they lack autonomy and independence in learning science; children experience 'fun' science activities that fail to deepen or develop new learning; children are not encouraged to use their own curiosity, scientific interests and questions in their science learning; children are engaged in prescriptive practical work that lacks purpose; children do not draw on their learning from prior scientific skills, they do not build on repeated and regular experiences; children rarely see themselves, their families, community members or their teachers as scientists; children do not apply literacy and numeracy skills in science at the standard they use in English and mathematics.

Bianchi *et al.* (2021) appreciated that their report was hard hitting and recognised that it would take time to address the issues in meaningful ways. The ambition for change included strategic leadership where changes made were underpinned by a clear rationale in order to improve the experiences that children encounter in their science learning. The authors made two key recommendations from their findings. The first was the continuation of monitoring children's experiences through a programme of school reviews with the purpose of gathering insights into the issues impacting on children's learning of science and the extent to which these priorities were being addressed. The second was an annual consultation of data through engagement with sector colleagues to discuss the key issues arising and how the sector could work collaboratively to mitigate them. The potential impact of this report can be seen in the clear identification of current issues affecting learning and the ways in which it invites ongoing consultation and research from a wide range of stakeholders.

In 2021, Ofsted published their first research review focusing on science education. The review included previously published research on factors that influence the quality of science education and not on their inspection findings. It sought to identify features associated with high-quality science education, recognised that there was not just one way of defining high-quality science education, and identified a number of principles that guide how inspectors look at what they see. It considered factors including why teach science, what to teach at what age, how to teach, how to assess progress, how to motivate, how to develop materials, how best to deploy the teaching force in school systems, and how to equip the teaching workforce through initial and ongoing development.

The Primary Science Education Consultancy (2021) noted that substantial sections of the report were of greater relevance to secondary than primary schools. They expanded on the key areas from the report that related to primary science in order to help schools consider the findings. These included: curriculum progression – what it means to get better at science; organising knowledge within the subject curriculum; other curricular considerations – including the most effective use of time available providing a variety of learning approaches; practical work; pedagogy; assessment; and systems and subjects at school level.

Leach (2021) critiqued Ofsted's review in the way that it presented research evidence about teaching and learning in a highly abstracted way and thus 'risks reinforcing the view that (science) education adds nothing to practitioner wisdom and common sense' (p.68). Although commending engagement with research evidence, Leach concluded that there is danger that ideology was prioritised over evidence in shaping teachers' practice in schools. He identified effective practice as not simply the adoption of a certain approach to teaching (such as dialogic) but how this approach was executed. In his response to Ofsted's review, Leach argued that it is important for teaching approaches to be designed into the activity according to the teaching purpose.

If the findings are to be taken seriously, Leach (2021) argued that fundamental change would be required to the ways in which science teachers, ITE, curriculum developers and assessment agencies carry out their work in science education, supported by the appropriate resourcing. Emerging from Leach's critique is the question of how research evidence can be meaningfully used by schools to change practice so that primary science can become more effective and perhaps ambitious. One approach to supporting and promoting research among practising teachers can be achieved through open access, evidence-based research journals, such as the Journal of Emergent Science, which is focused on primary aged learners.

There is a vast literature on primary science which can be organised in a number of ways. This review identified a range of factors that influence primary science which were organised into five themes: monitoring and assessment; curriculum design and progression; teaching approaches; societal, national and global influence; and teacher knowledge, development and support. Concerns associated with access and equity - amplified by the loss of learning through the COVID-19 pandemic - and with government claims of post-Brexit opportunities for the future of STEM (Department for Business, Energy & Industrial Strategy *et al.*, 2020), it is important to maintain a research focus on primary science. Whilst barriers to children's learning have been clearly identified, it is encouraging to note that recent data indicates an improving picture for primary science. The contextual factors identified throughout this review provide an understanding of the practice of primary science in

England since the introduction of the most recent curriculum and will be used to inform and refine the search terms for the systematic review.

Appendix 1A: Teaching materials listed in the review

Teaching materials included in the review above are listed in alphabetical order. This is not an exhaustive list of materials for teaching primary science.

BEST (Best Evidence Science Teaching). Developed by the University of York funded by the Salters' Institute. Although designed as a secondary resource, it is included here as data showed that approximately a quarter of the downloads of materials designed for the 11-14 year age group were by primary schools (Waller, 2021). 'The research-informed resources provide progression toolkits for key concepts; appropriately-sequenced steps for learning progression; diagnostic questions that provide evidence of learning and of common misunderstandings; and response activities that challenge misunderstandings, promote metacognition and encourage conceptual progression.' https://www.stem.org.uk/resources/collection/440721/best-evidence-science-teaching

Empiribox Primary School Science. A primary school science specialist provider. 'A complete service to help teachers do inspiring practical science with any primary school year group. We provide lesson plans, assessments and equipment.' https://empiribox.com/

Explorify. Funded by Wellcome as part of its UK-wide Primary Science Campaign. Designed for teachers, by teachers. 'Raise the profile of science in your school. Confidently teach science and inspire your pupils.' A range of activities are provided to 'inspire questioning, deepen thinking and extend reasoning skills.' https://explorify.uk/

Oak National Academy. Although the aims of teaching science are not clear from the website, references in more recent literature (e.g. Lough, 2021) were made to resources by Oak National Academy used during the COVID-19 pandemic. Part of the Reach Foundation and supported by the DfE, this resource was 'made by teachers for teachers' in April 2020 as a rapid response to the coronavirus outbreak. Resources include fully sequenced curriculum maps, formative assessment tools, individual lesson plans to edit or download, teacher-made lesson slides, videos, worksheets and quizzes. https://www.thenational.academy/

Primary Science Teaching Trust. A charity that aims to facilitate the development and dissemination of excellence in primary science. 'Our vision is to see excellent teaching of science in every primary classroom in the UK. Provides a wealth of resources to support teaching, learning, assessment.' https://pstt.org.uk/

Terrific Scientific. BBC drawing on 10 mass-participation science investigations linked to 'real' research undertaken by universities. 'A set of curriculum-linked primary science resources from BBC Learning for Key Stage 2 aimed at encouraging scientific enquiry.' https://www.bbc.co.uk/teach/terrific-scientific

Wow Science. Created by Primary Science Teaching Trust (PSTT) in collaboration with Learning Science Ltd. 'Searching out the best primary science activities. Provides links to the best primary science learning materials on the web, helping children to enjoy science both inside and outside the classroom.' https://wowscience.co.uk/

Appendix 1B: Primary science resources, CPD and networks

Primary science networks and resource and CPD providers included in the narrative review are listed in alphabetical order below. This is not an exhaustive list.

ASPIRES. Longitudinal research project studying young people's science and career aspirations currently in its third stage of funding, ASPIRES 3. ASPIRES and ASPIRES 2 were first based at King's College London, moving to the UCL Institute of Education in March 2017 during ASPIRES 2. https://www.ucl.ac.uk/ioe/departments-and-centres/departments/education-practice-and-society/aspires-research

The Association for Science Education (ASE) is the professional association for teachers of science and exists to improve the teaching of science. A Registered Charity with a Royal Charter, ASE is an active membership body that has been supporting all those involved in science education from pre-school to higher education. Provides a range of resources (for teaching and assessment) and journals including Primary Science (PS), a themed journal for all those involved in primary science education for children aged 3-12. https://www.ase.org.uk/primary-resources

Children Challenging Industry (CCI) at the Centre for Industry Education Collaboration (CIEC). Combining tailored training for industry partners and primary school staff with fully resourced problem-solving classroom activities and interactive site visits to industry. https://www.york.ac.uk/ciec/cci/

CREST. A scheme that inspires young people to think and behave like scientists and engineers. 'Inspire your primary-aged students with short, hands-on activities that challenge them to explore the world around them. Awards include Star (5-7 year-olds) and Superstar (7-11 year-olds). https://www.crestawards.org/

Journal of Emergent Science (JES) produced by the PSTT and ASE has a 'research focus on primary aged learners and we see it as a pivotal step in supporting and promoting research amongst practising teachers. By having open access to JES, teachers will be able to access evidence-based research that will inform their teaching practices.' https://www.ase.org.uk/resources/journal-of-emergent-science

Ogden Trust. In primary education the Trust aims to: Raise the profile of science in the primary curriculum; enhance confidence in the planning, delivery and assessment of primary science; and support science subject leaders in creating a sustainable network. Leads CPD. Builds School Partnerships and provides funding to support events, activities, trips and training. https://www.ogdentrust.com/

Primary Science Education Consultancy (PSEC). 'Our role is to unlock the potential in schools and teachers to inspire their pupils about science, while also developing their confidence in teaching and assessing their children's knowledge and skills.' Their work includes: setting up and leading regular science subject leader meetings; developing and delivering annual programmes of tailored CPD; devising and delivering specific projects; offering advice; providing links to a range of websites and resources. https://www.primary-science.co.uk/

Primary Science Capital Project 'Working with teachers to develop a science capital approach for primary schools.' Primary Science Capital Teaching Approach (PSCTA) 'a teaching framework that helps teachers to reflect on and develop new ways to promote children's engagement and identification with science.' IOE at UCL, King's College London, PSTT and the Ogden Trust.

https://www.ucl.ac.uk/ioe/departments-and-centres/departments/education-practice-and-society/stem-participation-social-justice-research/primary-science-capital-project

PSQM - Primary Science Quality Mark (PSQM) based at The University of Hertfordshire and supported by the PSTT. The PSQM is a national school improvement programme that raises the standard of teaching and learning in primary science through improving the practice of the science subject leader. It is delivered via the PSQM Hubleaders – a national network of in-house trained primary science training professionals. http://www.psqm.org.uk/

The Primary Science Teaching Trust (PSTT) is a charitable trust helping to improve the teaching and learning of primary science across the UK. Their strategy consists of three approaches: supporting award-winning primary science teachers through the Primary Science Teacher College; supporting groups of schools working together through the Cluster Programme; and supporting research and innovation through Academic Collaborators based at various universities. https://pstt.org.uk/

Reach Out CPD - Courses to develop teacher knowledge and confidence in order to inspire children. Resources including videos, activities and experiments are provided. Reach Out CPD is a partnership between Tigtag (part of Twig Education) and Imperial College London. https://www.reachoutcpd.com/

Science and Engineering Education Research and Innovation Hub (SEERIH) at The University of Manchester. Partners include the PSST and the Ogden Trust. SEERIH is 'passionate about defining quality science teaching and learning. We listen and respond to teachers, advocating, challenging and influencing national policy for science and engineering education.' Their work includes: professional learning programmes; establishment of primary science subject leader networks and community-led science clusters to facilitate inter-school collaboration with focus on supporting Year 6-7 transition; Deep Dive one-day intensive school reviews (e.g. Science across the City, Stoke-on-Trent); an annual Primary Science Conference; online resources to support learning; and research projects connecting with a range of stakeholders. https://www.seerih.manchester.ac.uk/

STEM Learning. 'If you're a primary teacher, we're here to support you at every stage of your career, and we share your dedication to raising young people's engagement and achievement in STEM.' Their work includes Resources, CPD, STEM Ambassadors, STEM Clubs. https://www.stem.org.uk/primary

The Teacher Assessment in Primary Science (TAPS) project is based at Bath Spa University and funded by the Primary Science Teaching Trust (PSTT). TAPS 'aims to develop support for a valid, reliable and manageable system of primary school science assessment which will have a positive impact on children's learning. The TAPS pyramid tool provides a structure to help schools evaluate and develop their assessment processes. The rich formative assessment information collected by teachers in the course of ongoing classroom work is also utilised for summative purposes.' https://www.bathspa.ac.uk/projects/taps/

Thinking, Doing, Talking Science (TDTS) is an interactive four-day training programme for teachers that focuses on developing creative and challenging science lessons that encourage pupils to use higher-order thinking skills. 'Focuses on developing creative and challenging science lessons that encourage pupils to use higher-order thinking skills. TDTS teachers enable their pupils to think and talk about scientific concepts through dedicated discussion times, they provide them with a wide range of opportunities for creative investigations and problem solving and they focus pupils' recording so that there is always time for practical science.' In partnership with Science Oxford and Oxford Brookes University. Supported by the EEF and PSTT. https://tdts.org.uk/

The University of York Science Education Group (UYSEG) 'aims to make a sustained positive impact on the outcomes of both formal and informal science education through: high-quality research that has an impact on policy and practice; the development and evaluation of research-informed curricula that illustrate the importance of science; the training and support of practitioners.' Their work includes: Best Evidence Science Teaching (BEST) is a major research-informed curriculum development project, which aims to transform science education research into practise by making research-informed, open-access resources freely available to science teachers (written for age 11-16 and used by primary schools); the BEST Primary Pilot project; Year 6-7 transition work in Stoke-on-Trent (Science Across the City and DfE); ASE BEST Bites CPD project; leadership of a range of research projects. https://www.york.ac.uk/education/research/uyseg/

Wellcome is a global charitable foundation. UK-wide Primary Science Campaign with the vision that 'all pupils experience an exciting, inspiring and relevant science education at primary school that leaves them well-prepared to progress further in science, and well-informed about science in their everyday lives.' Funders of research across disciplines. https://wellcome.org/

References

Abrahams, I., Reiss, M., & Sharpe, R. (2014). 'The impact of "Getting practical work in science" continuing professional development programme on teachers' ideas and practice in science practical work', *Research in Science and Technological Education, 32*(3), 263–280.

Archer, L., Osborne, J., DeWitt, J., Dillon, J., Wong, B., & Willis, B. (2013a). *ASPIRES: Young people's science and career aspirations, age 10-14*. King's College London.

Archer L., DeWitt J., Osborne J., Dillon, J., Willis, B., & Wong, B. (2013b). 'Not girly, not sexy, not glamorous': primary school girls' and parents' constructions of science aspirations. *Pedagogy Culture and Society, 21*(1), 171–194. https://doi.org/10.1080/14681366.2012.748676

Archer L., Moote J., Macleod E., Francis, B., & DeWitt, J. (2020). *ASPIRES 2: Young people's science and career aspirations, age 10–19*. London, UK: UCL Institute of Education. https://discovery.ucl.ac.uk/id/eprint/10092041/

ASE (2016) A response to the House of Commons Education Committee's inquiry on Primary Assessment https://www.ase.org.uk/sites/default/files/ASE%20response%20HoC%20Education%20Inquiry%20on%20Primary%20Assessment%20%28Oct2016%29.pdf

Assessment Reform Group. (2002) *Assessment for Learning: 10 Principles. Research-based principles to guide classroom practice Assessment for Learning.* https://www.storre.stir.ac.uk/bitstream/1893/32458/1/Assessment%20for%20learning%2010%20principles%202002.pdf

Barnes, J. (2015). *Cross-curricular learning 3-14*. 3rd ed.Sage.

Bennett, J., Hanley, P., Abrahams, I., Elliott, L., & Turkenburg-van Diepen, M. (2019) Mixed methods, mixed outcomes? Combining an RCT and case studies to research the impact of a training programme for primary school science teachers. *International journal of science education, 41*(4), 490–509. Routledge.

Beverley, N. Lawrence, L., Turford, B. and Turner, J. (2016). Snap Science Assessment Years 1 - 6. Collins. https://collins.co.uk/products/9780008199333

Bianchi, L. (2015) Being right isn't persuasive. *Primary Science, 137*, 15-17. https://seerih-innovations.org/media/doc/Being_right_isnt_persuasive_BIANCHI_2015.pdf

Bianchi, L. (2021). Great Science Share for Schools. Purpose and Vision. https://www.greatscienceshare.org/aboutus

Bianchi, L., Whittaker, C., & Poole, A. (2021). *The 10 key issues with children's learning in primary science in England.* https://www.scienceacrossthecity.co.uk/wp-content/uploads/2021/03/3634_Childrens_Learning_in_Primary_Science_Report_2020_v8.pdf

Black, P., & Wiliam, D. (1998). *Inside the black box - raising standards through classroom assessment.* GL Assessment.

Blackmore, K., Howard, C., & Kington, A. (2018). Trainee teachers' experience of primary science teaching, and the perceived impact on their developing professional identity. *European journal of teacher education, 41*(4), 529–548. https://doi.org/10.1080/02619768.2018.1462330

Bloom, A. (2017). Subjects other than English and maths barely feature in Ofsted reports. Tes findings highlight concerns about the 'narrowing' school curriculum. https://www.tes.com/news/exclusive-subjects-other-english-and-maths-barely-feature-ofsted-reports

Bloom, B. S. (1956). *Taxonomy of educational objectives, Handbook I: The cognitive domain*. David McKay Co Inc.

Borthwick, A., & Cross, A. (2015). Primary pupils: working scientifically or working mathematically? *Journal of Educational & Instructional Studies in the World, 5*(4), 75-84.

Bostrom, C. (2016). Experiencing Child-Led Science in Science Week. *Primary Science, 142*, 31-33. Association for Science Education.

British Science Association (2023). British Science Week (britishscienceweek.org)

Canovan, C., & Fallon, N. (2021). 'It isn't a priority.' Will primary science learning loss be forgotten? *The Journal of Emergent Science, 21*, 46-53. Association for Science Education. http://clok.uclan.ac.uk/38472/

Chowdhuri, N., King, M., & Archer, L. (2021). *The Primary Science Capital Teaching Approach: teacher handbook*. UCL Institute of Education: London, UK. https://www.ucl.ac.uk/ioe/departments-and-centres/departments/education-practice-and-society/stem-participation-social-justice-research/primary-science-capital-project

Clarke, S. (2001). *Unlocking Formative Assessment Practical Strategies for Enhancing Pupils' Learning in the Primary Classroom.* Abingdon Hodder Education.

Confederation of British Industry, CBI. (2015). *Tomorrow's World: inspiring primary scientists*. https://www.stem.org.uk/resources/elibrary/resource/35987/tomorrows-world-inspiring-primary-scientists

Crompton, Z. (2013). What Is Science? Some Research from Primary Schools. *Education in science: the bulletin of the Association for Science Education*. Association for Science Education. 253, 12-13.

Cross, A., & Board, J. (2014). *Creative Ways to Teach Primary Science.* Milton Keynes: Open University Press

Cutting, R., & Kelly, O. (2014). *Creative Teaching in Primary Science*. SAGE.

Danielsson A. T., & Warwick P. (2014). 'You Have to Give Them Some Science Facts': Primary Student Teachers' Early Negotiations of Teacher Identities in the Intersections of Discourses About Science Teaching and About Primary Teaching. *Research in science education, 44*(2), 289-305.

Davenport, C., & Shimwell, J. (2019). Careers Advice and Changing Stereotypes in the Primary Classroom. *Primary Science, 157*, 29-30.

Davies, D., Collier, C., Earle, S., Howe, A. & McMahon, K. (2014). Approaches to Science Assessment in English Primary Schools. *Teachers' Summary. Centre for Research in Early Scientific Learning (CRESL), Bath Spa University*. https://pstt.org.uk/application/files/6514/5761/9873/Approaches_to_science_asst_Full_report_Oct14.pdf

Davies, D. J., Earle, S., McMahon, K., Howe, A. & Collier, C. (2017). Development and exemplification of a model for Teacher Assessment in Primary Science. *International journal of science education, 39*(14), 1869–1890. https://doi.org/10.1080/09500693.2017.1356942 Department for Business, Energy & Industrial Strategy, Department for Education, UK Research and Innovation, The Rt Hon Sir Alok Sharma KCMG MP, &The Rt Hon Sir Gavin

Williamson CBE MP (2020). *Multi-million government investment in the future of UK science*. Press release. https://www.gov.uk/government/news/multi-million-government-investment-in-the-future-of-uk-science

Department of Education and Science and the Welsh Office (1988). *National Curriculum Task Group on Assessment and Testing: a report* http://www.educationengland.org.uk/documents/pdfs/1988-TGAT-report.pdf

Department for Education. (2012, June 11). *New primary curriculum to bring higher standards in English, maths and science.* https://www.gov.uk/government/news/new-primary-curriculum-to-bring-higher-standards-in-english-maths-and-science

Department for Education (2013) *The national curriculum in England: key stages 1 and 2 framework document.* https://www.gov.uk/government/publications/national-curriculum-in-england-primary-curriculum

Department for Education (2019a). *Education inspection framework.* https://www.gov.uk/government/collections/education-inspection-framework

Department for Education (2019b). *Early career framework. A framework of standards to help early career teachers succeed at the start of their careers.* https://www.gov.uk/government/publications/early-career-framework

Department for Education. (2020). Key stage 2: science sampling tests https://www.gov.uk/government/publications/key-stage-2-science-sampling-tests

Department for Education (2022). *Sustainability & climate change: A strategy for the education & children's services systems.* https://www.gov.uk/government/publications/sustainability-and-climate-change-strategy/sustainability-and-climate-change-a-strategy-for-the-education-and-childrens-services-systems

Department for Education and Employment (1999). *The National Curriculum Handbook for primary teachers in England Key stages 1 and 2.* http://www.educationengland.org.uk/documents/pdfs/1999-nc-primary-handbook.pdf

Dunlop, L., Airey, J., Turkenburg-van Diepen, M., & Bennett, J. (2020). Close encounters between young people and human spaceflight. *International Journal of Science Education Part B, 10*(3), 187–203. https://doi.org/10.1080/21548455.2020.1730018

Earle, S. (2014). Formative and summative assessment of science in English primary schools: evidence from the Primary Science Quality Mark. *Research in science & technological education, 32*(2), 216–228. https://doi.org/10.1080/02635143.2014.913129

Earle, S. (2015). An exploration of whole-school assessment systems. *Primary Science, 136*, 20-22. Association for Science Education. http://researchspace.bathspa.ac.uk/5227/1/Earle%20Primary%20Science%20Final.pdf.

Earle, S., Mc Mahon, K., Collier, C., Howe, A., & Davies, D. (2017). The Teacher Assessment in Primary Science (TAPS) school self-evaluation tool. https://pstt.org.uk/application/files/1614/8768/8133/TAPS_teacher_summary_v3_Jan_2017_to_print.pdf

Earle, S. & Turner, J. (2020). What has happened to teacher assessment of science in English primary schools? Revisiting evidence from the Primary Science Quality Mark. *Research in Science & Technological Education*. Routledge: 1–17. https://doi.org/10.1080/02635143.2020.1834375

Eco-schools Green Flag Award. https://www.eco-schools.org.uk/about/eco-schools-green-flag-other-awards/

Eshach, H., & Fried, M. N. (2005). Should science be taught in early childhood? *Journal of Science Education and Technology, 14*(3), 315–336.

George, M. (2018, March 23). *Why is primary science dying?* The Times Educational Supplement 5292, London (Mar 23, 2018). https://www.tes.com/magazine/article/why-primary-science-dying

Grant, M. J., & Booth, A. (2009). A typology of reviews: an analysis of 14 review types and associated methodologies. *Health Information and Libraries Journal*, *26*(2), 91–108.

Hanley, P., Wilson, H., Holligan, B., & Elliott, L. (2020). Thinking, doing, talking science: the effect on attainment and attitudes of a professional development programme to provide cognitively challenging primary science lessons. *International journal of science education, 42*(15), 2554–2573. https://doi.org/10.1080/09500693.2020.1821931

Harlen, W. (2013). Inquiry-based learning in science and mathematics. *Review of science, mathematics and ICT education*, *7*(2), 9-33.

Harlen, W. (2015). Towards big ideas of science education. *School Science Review, 97*(359), 97–107. Association for Science Education. https://www.ase.org.uk/resources/school-science-review/issue-359/towards-big-ideas-science-education

Harlen, W., Bell, D., Cutler, M., Goldsworthy, A., Hall, A., Harrison, C., Howard, S., Keogh, B., Lawrence, L., Naylor, S., Reiss, M., Shallcross, D., & Turner, J. (2012). *Developing policy, principles and practice in primary school science assessment.* Nuffield Foundation. https://www.nuffieldfoundation.org/wp-content/uploads/2019/11/Developing_policy_principles_and_practice_in_primary_school_science_assessment_Nuffield_Foundation_v_FINAL.pdf

Harlen, W., Bell, D., Devés, R., Dyasi, H., Fernández de la Garza, G., Léna, P., Millar, R., Reiss, M., Rowell, P., & Yu, W. (2010). *Principles and big ideas of science education.* Harlen, W (ed) https://www.ase.org.uk/bigideas

Harlen, W., Bell, D., Devés, R., Dyasi, H., Fernández de la Garza, G., Léna, P., Millar, R., Reiss, M., Rowell, P., & Yu, W. (2015). *Working with Big Ideas in Science Education.* https://www.ase.org.uk/download/file/fid/6740

Harlen, W. & Qualter, A. (2015). *The Teaching of Science in Primary Schools.* Abingdon: Routledge.

Hartmeyer, R., Stevenson, M. P., & Bentsen, P. (2018). A systematic review of concept mapping-based formative assessment processes in primary and secondary science education. *Assessment in Education: Principles, Policy & Practice, 25*(6), 598–619.

Hashweh, M. (1985). The exploratory study of teacher knowledge and teaching: The effects of science teachers' knowledge of subject matter and their conceptions of learning on their teaching [PhD]. Stanford University.

Heathcote, D. & Bolton, G.M. (1995). *Drama for Learning: Dorothy Heathcote's Mantle of the Expert Approach to Education.* London: Pearson.

Hobbs, L., Stevens, C., Hartley, J., Ashby, M., Lea, I., Bowden, L., Bibby, J., Jackson, B. W., McLaughlin, R., & Burker, T. (2019). Using Minecraft to engage children with science at public events. *Research for All, 3*(2), 142-160. UCL Press. https://doi.org/10.18546/RFA.03.2.03

Hodgson, C. (2010). *Assessment for learning in primary science: practices and benefits.* NFER. https://www.nfer.ac.uk/publications/aas02/aas02.pdf

Hopwood-Stephens, I. (2018) Teachers' attempts to improve assessment practice in primary science are influenced by job role and teaching experience. *JES, 15*(Summer), 57-66.

House of Commons Science and Technology Select Committee (2017). Science communication and engagement. https://publications.parliament.uk/pa/cm201617/cmselect/cmsctech/162/16207.htm

Keeble, R. (2016) *Effective Primary Teaching Practice 2016.* Teaching Schools Council's Report. https://tactyc.org.uk/wp-content/uploads/2016/08/Effective-primary-teaching-practice-2016-report-web.pdf

Lawrence, L., Beverley, N. & Turford, B. (2016). *Snap Science Assessment Years 1-6* (J. Turner (ed.)). Collins.

Leach, J. (2021). Research evidence, policy and practice: reflections on the Ofsted research review on science. *School Science Review, 103*(382), 67–72. https://research.leedstrinity.ac.uk/en/publications/research-evidence-policy-and-practice(d3a4e635-b025-48f6-8551-a4e52e32aa71).html

Leonardi, S., Lamb, H., Howe, P., & Choudhury, A. (2017). *"State of the nation" report of UK primary science education: baseline research for the Wellcome Trust Primary Science Campaign*. Wellcome Trust. https://wellcome.ac.uk/sites/default/files/state-of-the-nation-report-of-uk-science-education.pdf

Leonardi, S., Lamb, H. Harrison, J., Steer, R., & Howe, P. (2019). *What pupils think of science in primary schools. Baseline research for the Wellcome Primary Science Campaign.* https://wellcome.org/sites/default/files/what-pupils-think-of-science-in-primary-schools.pdf

Lister, J. (2018). Q&A Jenny Lister. *Astronomy & Geophysics, 59*(4), 4.43–4.43. https://doi.org/10.1093/astrogeo/aty196

Lough, C. (2021). Covid: How remote learning changed during pandemic. Academies used Oak National more than other state schools, and four other key findings from the online academy's data. *TES November 2021*. https://www.tes.com/news/covid-how-remote-learning-changed-during-pandemic-online-learning

Mackintosh, J., White, E., & Dickerson, C. (2017). Developing teachers as leaders of science in primary schools. *Journal of Emergent Science, 12*(Winter 2016/2017), 64-71. Association for Science Education (ASE). http://uhra.herts.ac.uk/handle/2299/19062

Markwick, A., & Clark, K. (2016). Science + maths=a better understanding of science! *Primary Science, 145*, 5–8. Association for Science Education.

Markwick, A. (2019). Working scientifically and assessing mathematics mastery. *Primary Science, 158*, 25-27. Association for Science Education. https://discovery.ucl.ac.uk/id/eprint/10114803/1/PS%20158%20Working%20scientifically%20and%20assessing%20mathematics%20mastery.pdf.

Maxwell, D., Pillatt, T., Edwards, L., & Newman, R. (2019). Applying Design Fiction in Primary Schools to Explore Environmental Challenges. *The Design Journal, 22*(1), 1481–1497. https://doi.org/10.1080/14606925.2019.1594972

McCrory, A. (2017). Scientific enquiry and engaging primary-aged children in science lessons: reflections on provision for scientific enquiry in primary schools in England and Wales: Part 1. *Journal of Emergent Science, 13*(Summer 2017), 5-9. https://discovery.ucl.ac.uk/id/eprint/1574818/

McCullagh, J., & Doherty, A. (2018). Digital Makeover: What do pre-service teachers learn from microteaching primary science and how does an online video analysis tool enhance learning? *Teacher Education Advancement Network Journal, 10*(2), 15–28. http://ojs.cumbria.ac.uk/index.php/TEAN/article/view/473

McGregor, D. (2014). Chronicling innovative learning in primary classrooms: conceptualizing a theatrical pedagogy to successfully engage young children learning science. *Pedagogies: An International Journal, 9*(3), 216–232. https://doi.org/10.1080/1554480X.2014.899544

McGregor, D., Baskerville, D., Anderson, D., & Duggan, A. (2019). Examining the use of drama to develop epistemological understanding about the nature of science: a collective case from experience in New Zealand and England. *International journal of science education. Part B. Communication and public engagement, 9*(2), 171–194. https://doi.org/10.1080/21548455.2019.1585994

Mepsted, J. (2018). Supporting the Teaching and Assessment of Working Scientifically. *Primary Science, 151*, 19-20. https://www.ase.org.uk/journals/primary-science/2018/03/2018/

Naylor, S. & Keogh, B. (2015). Talking and thinking using concept cartoons. What have we learnt? *School Science Review, 97*(359), 61-67.

Nunes, T., Bryant, P., Strand, S., Hillier, J., Barros, R. & Miller-Friedmann, J. (2017). *Review of SES and Science Learning in Formal Educational Settings.* EEF and Royal Society. https://royalsociety.org/~/media/policy/topics/education-skills/education-research/evidence-review-eef-royalsociety-22-09-2017.pdf?la=en-GB

Ofsted (2015). Key stage 3: the wasted years? https://www.gov.uk/government/publications/key-stage-3-the-wasted-years
Ofsted. (2020). COVID-19 series: briefing on schools; Evidence from pilot visits to schools between 29 September and 23 October 2020. Ofsted. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/933490/COVID-19_series_briefing_on_schools_October_2020.pdf
Ofsted (2021). Research review series: science. https://www.gov.uk/government/publications/research-review-series-science/research-review-series-science
Oliver, A. (2013). *Creative Teaching: Science in the Early Years and Primary Classroom.* London: Fulton.
Orchard, K., Hall, B., & Milward, Z. (2019). A Whole-School Vision for Science. *Primary Science, 159*, 5-8. Association for Science Education. https://www.ase.org.uk/resources/primary-science/issue-1590/whole-school-vision-science
Parry, S., Britain, E. & Allen, M. (2019). Where has all the science gone? *Primary science, 157*, 14–15. https://eprints.kingston.ac.uk/id/eprint/43146/1/Allen-M-43146-VoR.pdf
Primary Science Education Consultancy, THE Partnership, Association for Science Education. (n.d.). PLAN primary science assessment resources. https://www.planassessment.com/
Primary Science Education Consultancy (2021). *Thoughts on Ofsted's research review of the factors that influence the quality of science education.* Blog. https://www.primary-science.co.uk/post/thoughts-on-ofsted-s-research-review-of-the-factors-that-influence-the-quality-of-science-education
Primary Science Teaching Trust (2021). *Our Vision.* https://pstt.org.uk/who-we-are/vision
Qualifications and Curriculum Agency (QCA) (established in 1997 and dissolved in 2011). *Schemes of Work. Key Stages 1and 2: the units.* https://webarchive.nationalarchives.gov.uk/ukgwa/20100607215759/http://www.standards.dfes.gov.uk/schemes2/science/?view=get
Quigley, C. F., & Herro, D. (2016). Finding the Joy in the Unknown: Implementation of STEAM Teaching Practices in Middle School Science and Math Classrooms. *J Sci Educ Technol, 25*, 410–426. https://doi.org/10.1007/s10956-016-9602-z
Redhead, K. (2021). Whole school curriculum planning for primary science...post pandemic. PSTT. https://pstt.org.uk/resources/curriculum-materials/Science-Curriculum-Coverage
Ridley, K. (2014). Young 'science ambassadors' raise the profile of science. *Primary science, 133*, 14-16. Association for Science Education. http://www.ase.org.uk/journals/primary-science/2014/05/133/
Roberts, D. (2019). The challenges facing primary science education. *Headteacher update.* https://www.headteacher-update.com/best-practice-article/the-challenges-facing-primary-science-education/213115
Robertson, A. (2017, September 30). *Only 1 in 4 pupils meets expected standard in science tests*. https://schoolsweek.co.uk/only-1-in-4-pupils-meets-expected-standard-in-science-tests/
Roche, J., Bell, L., Galvão, C., Golumbic, Y. N., Kloetzer, L., Knoben, N., Laakso, M., Lorke, J., Mannion, G., Massetti, L., Mauchline, A., Pata, K., Ruck, A., Taraba, P., & Winter, S. (2020). *Citizen Science, Education, and Learning: Challenges and Opportunities.* https://doi.org/10.3389/fsoc.2020.613814
Rogers, B. (2021). Primary science – How and why Paradigm Trust overhauled its science curriculum. Teachwire. https://www.teachwire.net/news/primary-science-how-and-why-paradigm-trust-overhauled-its-science-curriculum
Royal Society (2021). *A broad and balanced curriculum.* https://royalsociety.org/topics-policy/projects/vision/science-to-18/#:~:text=The%20Vision%3A%20All%20young%20people,18%2C%20including%20maths%

20and%20science.&text=Inspirational%20mathematics%20and%20science%20lessons,practical%20work%20and%20problem%20solving.
Royal Society & the British Academy (2018). *Harnessing educational research.* https://royalsociety.org/-/media/policy/projects/rs-ba-educational-research/educational-research-report.pdf Related webpage: https://royalsociety.org/topics-policy/projects/vision/evidence-in-education/
Royal Society & CBI (2016). *Making education your business: A practical guide to supporting STEM teaching in schools and colleges.* Issued: May 2016 DES4026 https://royalsociety.org/~/media/policy/Publications/2016/business-toolkit.pdf
Rushton, E. A. C., & Reiss, M. J. (2021). Middle and high school science teacher identity considered through the lens of the social identity approach: a systematic review of the literature. *Studies in Science Education, 57*(2), 141–203. https://doi.org/10.1080/03057267.2020.1799621
Smith, K. (2016). *Working Scientifically: A Guide for Primary Science Teachers*. Routledge.
Stagg, B. C. (2020). Meeting Linnaeus: improving comprehension of biological classification and attitudes to plants using drama in primary science education. *Research in science & technological education, 38*(3), 253–271. https://doi.org/10.1080/02635143.2019.1605347
STEM (2021). Tim Peake Project Activities. https://www.stem.org.uk/resources/community/collection/22312/tim-peake-project-activities
Stubberfield, L. (2021). Slowly does it - using research to improve primary science. *Journal of Emergent Science, 20*, 5-10. https://www.ase.org.uk/system/files/Stubberfield.pdf.
Trew, A. J., Early, C., Ellis, R., Nash, J., Pemberton, K., Tyler, P., Skerry, C., Bird, L., Shallcross, K. R., Harrison, T. G., & Shallcross, D. E. (2023). Can current science research in the biological sciences be used in primary school children's scientific enquiry? *Journal of biological education, 57*(3), 455-468. https://doi.org/10.1080/00219266.2021.1924229
Turner, J. (2016). Primary Science Quality Mark--2016 Update. *Primary Science, 143*, 20-21. https://www.ase.org.uk/resources/primary-science/issue-143/primary-science-quality-mark-2016-update
Tyler, P. (2018). It's Not Just Games: Exploring the Science of Sport. *Primary Science, 152*, 11-15. https://www.ase.org.uk/journals/primary-science/2018/03/152/
Tyler, P. (2019). Turn your primary pupils into a real-life science research team (sponsored). How can you get pupils excited about science? Citizen science projects inspire and engage them in real-world exploration. https://www.tes.com/news/turn-your-primary-pupils-real-life-science-research-team-sponsored
Waller, N. (2021). What is BEST for primary science? *Primary Science, 169*, 26-29. https://www.ase.org.uk/resources/primary-science/issue-1690/what-best-primary-science-open-access
Ward, H. & Roden, J. (2016). *Teaching Science in the Primary Classroom*, 3rd ed. London: Sage.
Warren, C. (2021). An evaluation of the Science across the City project as seen through the lens of PSQM. Second report. https://www.scienceacrossthecity.co.uk/evaluators-feedback-2021/
Wellcome Trust. (2014). *Primary science: is it missing out?: recommendations for reviewing primary science*. Wellcome Trust. https://wellcome.org/sites/default/files/primary-science-is-it-missing-out-wellcome-sep14.pdf
Wellcome Trust (2015). *Review of the extent to which Ofsted reports mention science.* https://cms.wellcome.org/sites/default/files/extent-to-which-ofsted-reports-mention-science-wellcome-sep15.pdf
Wellcome Trust (2017). *A Review of Ofsted Inspection Reports: Science.* https://cms.wellcome.org/sites/default/files/review-of-ofsted-inspection-reports-2017.pdf
Wellcome Trust (2018). *Review of Ofsted school inspection reports 2017 to 2018*. https://wellcome.org/reports/review-ofsted-school-inspection-reports-2017-to-2018

Wellcome Trust (2020). *Young people's views on science education Science Education Tracker 2019 Wave 2.* https://wellcome.org/sites/default/files/science-education-tracker-2019.pdf

Wellcome Trust (2021) Primary science. https://wellcome.org/what-we-do/our-work/transforming-primary-science

Wilkinson, D. & Stallard, W. (2019). *Mastery in Primary Science*. SAGE.

Wilson, H. Holligan, B., & Hanley, P. (2018). Thinking, Doing, Talking Science. *Primary Science, 154*, 15-18. Association for Science Education.

Woodland Trust's Eco Green Tree School Award (2021). https://www.woodlandtrust.org.uk/support-us/act/your-school/green-tree-schools-award/?gclid=Cj0KCQiA7oyNBhDiARIsADtGRZbj-8E8Dgj9060I6-bhXYkpb1VYcAdLvGg1bjvCRMP2pMJRPpEHKsYaAoGIEALw_wcB&gclsrc=aw.ds

WWF's Sustainable Schools Award (2021). https://www.wwf.org.uk/get-involved/schools/tes-sustainable-schools-award

Appendix 2: Review of practice

Report on focus groups held with teachers of primary science in England

November 2021

Introduction and scoping

The purpose of the research on practice was to inform a systematic review of primary science teaching in two main respects: (i) to identify the outcomes of primary science education in England that teachers prioritise and (ii) to provide contextual information to ensure that the review is relevant to teachers' needs. It includes a narrative literature review (Appendix 1) to identify current practice in primary science teaching in England and the findings of online focus groups with teachers.

The focus of the direct interaction with teachers is to elicit teachers' own priorities, not limited to those mandated through the National Curriculum (NC). The NC objectives do not necessarily fully reflect what is intended and enacted by primary science teachers.

The questions this empirical research element of the review addresses are:

- What outcomes are important to teachers of primary science?
- What approaches are currently used?
- What are the influences on the teaching of primary science?

Methods

Participants

A total of 31 teachers took part in focus groups, from 27 schools, all with mixed intakes. The majority of teachers were science subject leads, or had held the post in the past. Tables 1 - 5 present the sample by school characteristics using the most recent data available from the Department for Education and the EEF Families of Schools database.

The schools represented by teachers in the sample have an above average percentage of children who reached the expected standard in reading, writing and maths (combined). Where a judgement was available, the majority of schools in the sample had Good or Outstanding Ofsted judgements.

Teachers participated anonymously and detailed demographics were not collected. Findings must be interpreted cautiously given the over-representation of teachers from schools with: a below average percentage of: children who are eligible for free school meals (FSM); children who speak English as an Additional Language (EAL); or children with special educational needs (SEN) with a statement or education and healthcare plan.

Ofsted judgement	Requires improvement	Good	Outstanding	Not available (NA)
Number of schools	1	14	6	6

Table 1: Most recent Ofsted judgements on schools represented by teachers in the sample. In 2019, 86% schools had been judged to be Good or Outstanding.

Pupils eligible for FSM (%)	0-10	11-20	21-30	31-40	41-50	51-60	61-70
Number of schools	8	9	5	3	0	1	1

Table 2: Most recent percentages of children eligible for free school meals (FSM) in schools represented by teachers in the sample. Total percentage of students in England eligible for FSM was 20.8% (January 2021).

Children with EAL (%)	0-10	11-20	21-30	31-40	41-50	51-60	61-70	NA
Number of schools	16	2	2	0	0	0	1	6

Table 3: Percentage of children with English as an additional language (EAL) in schools represented by the teachers in the sample (data from 2019). The total percentage of students with EAL in England was 21%.

Children with SEN	Below 3%	At or above 3%	Not available
Number of schools	16	5	6

Table 4: Percentage of children with SEN (as defined by having EHC plans/Statements of SEN) in schools represented by the teachers in the sample (data from 2019). In 2019, 3% of children in England had SEN (as defined by having EHC plans/Statements of SEN).

Pupils reaching expected standard (%)	40-50	51-60	61-70	71-80	81-90	91-100	NA
Number of schools	1	2	8	6	5	2	3

Table 5: Percentage of children meeting the expected standard at KS2 in Mathematics, Reading and Writing (data from 2019). In 2019, 65% of pupils reached the expected standard.

Approach

An interpretivist approach was used to find out what primary teachers think and do in relation to primary science education, what their challenges are, and how they deal with them. Interpretivism is guided by the research subject and allows the researcher to explore the subjects' understanding of the meaning of events and phenomena in their environment, commonly through qualitative research data (Brundrett & Rhodes, 2013). A total of eight online focus groups were held with primary teachers, science leads and headteachers, each lasting one hour. Ethical approval was granted by the University of York Department of Education Ethics Committee.

A total of 31 teachers were included, with the majority in the role of science subject lead. The focus groups explored teachers' priorities, what they saw as the aims of primary science, their anticipated and intended outcomes for primary science education, their approaches to primary science teaching, and the influences on their practice. The focus group guide can be found in the associated protocol for the research.

Focus groups were audio-recorded, transcribed and analysed in NVivo. Focus group contributions were anonymised upon transcription to preserve confidentiality of teacher participants. Responses have not been quantified, as is common practice in qualitative data analysis: we are interested in the full range of perceptions and approaches rather than the frequency of approaches (focus groups would not be the most appropriate way to answer quantitative questions).

A primarily deductive approach to qualitative analysis was used in the analysis of outcomes, where in the first round of coding, data was coded according to outcomes desired, using the framework developed by James and Brown (2007) which presents a typology of learning outcomes. This describes seven types of learning outcome: attainment, understanding, cognitive and creative, using, higher-order learning, dispositions, and membership, inclusion and self-worth. Two researchers coded the transcripts independently, discussing how to code outcomes which had links across categories. The final approach to coding is documented in Table 6. Whilst there may be some overlap between outcomes, we classified according to the best fit.

A second round of deductive coding focused on practice in primary science. The transcripts were coded again, this time focusing on the following categories of response: approaches, influences (barriers, enablers, experiences of COVID-19, research evidence) and experiences of subject leadership. On coding the ‘approaches’ category, an initial analysis coded according to type of approach (to assessment, curriculum, pedagogy or professional development), then codes were applied from the EEF Secondary Science Guidance report (EEF, 2018): self-regulation, preconceptions, practical work, modelling, memory, language of science and feedback. Additional approaches were coded inductively. Whilst these approaches may or may not be relevant to primary science, we coded all approaches and this framework helps to identify approaches common to both primary and secondary science. Similarly, an inductive approach was used within the categories of barriers, evidence, experiences of COVID-19 and leadership.

Findings

The findings from the focus groups reporting on primary science teaching are organised into the following three sections: outcomes, approaches and influences. ‘Influences’ includes reported barriers and enablers to primary science education, responses to the educational impact of the COVID-19 pandemic, leadership and professional development in primary science, and use of research evidence.

Outcomes valued by teachers

We found outcomes of importance for teachers across the seven types identified by James and Brown (2007). Teachers often cited more than one outcome of interest, and identified outcomes which could be categorised as more than one type. All focus groups identified outcomes related to dispositions, whereas the other outcomes were found in at least one focus group. We avoid quantification because it is the range of outcomes that we are interested in, not how popular they are.

All outcomes were classified using James and Brown’s typology (using the interpretation in Table 6); no additional outcome types were identified. One of the challenges identified by teachers was assessment of different desirable outcomes. Knowledge and understanding was perceived as easier to measure than other desirable outcomes such as the development of skills, curiosity or ‘thinking like a scientist’.

Outcome type	Definition (James and Brown, 2007)	Primary science interpretation
Attainments—often school curriculum based (literacy, numeracy, science).	These tend to be outcomes with relatively precise and familiar meanings, often implying mastery of specific rules or mental procedures associated with particular tasks, and are frequently open to straightforward traditional means of assessment.	We coded to this type of outcome where teachers mentioned attainment, coverage of the national curriculum or testing e.g. in preparation for learning at the next educational stage.
Understanding—of ideas, concepts, processes.	These imply more conceptual ways of thinking and making intelligent judgements about meanings. Rather than focusing on mastery of specific knowledge or procedures, understandings concern the development of reasoning and inference.	We coded for 'understanding' of any subject - not just science - and included understanding of both concepts and skills. Understanding vocabulary and thinking critically about concepts was included in this code.
Cognitive and creative—imaginative construction of meaning.	These focus on the capability to create or discover 'new' knowledge, whether personal, cognitive, aesthetic or practical. Here there is direct concern with knowledge innovation by learners themselves.	This has been interpreted as children being given opportunities to discover new knowledge for themselves and to determine their own methods of investigation.
Using—how to practise, manipulate, behave, engage in processes or systems.	This implies the development of known practical and technological skills and being able to apply them in appropriate contexts.	We coded data as 'using' where teachers mentioned application or practice of science or scientific skills in new contexts (including those beyond science).
Higher-order learning — advanced thinking, reasoning, metacognition.	These concepts of learning transcend other learning outcomes. They emphasise more sophisticated cognitive processes. Sometimes they lack a clarity of meaning, and agreement about how they are to be assessed is even more difficult to achieve.	We coded 'thinking like a scientist' here as distinct from specific examples of skills. Also references to the nature and culture of science.
Dispositions—attitudes, perceptions, motivations.	These reflect the affective conditions within learners that it is assumed will be necessary if they are to feel inclined to learn effectively in the context and community where the learning takes place	This was interpreted to include attitudes towards science, discussions about development of character and individual responses to science.
Membership, inclusion, self-worth	These reflect the learners' affinity towards, readiness to participate in and sense of worthwhile contribution to the group where the learning takes place. It is concerned with establishing social and substantive identities in ways that reinforce the inclination to learn.	We coded cultural and science capital here, and references to diversity and representation in science. This was interpreted as how children relate to the field, as opposed to the development of individual characteristics.

Table 6: Definition of outcome types, from James and Brown (2007)

We report examples of outcomes in each of the categories below, in alphabetical order. The section concludes with a table of keywords taken from across the outcomes, which will be integrated into searches during the systematic review.

Attainment

Attainment tended to be seen as a limited outcome of primary science. It was described as important for teacher planning and was rarely discussed as an end in itself.

> *Attainment...you obviously want to know what the children have got, what they've learned, what they've remembered, and you want to ensure that they remember from the previous year in order to move on to the next year, you know what I mean, they need to know what's come before so that they can access what they're doing now.*

Teachers described measures of attainment as important in ensuring that children were covering the curriculum, making progress and were being prepared for secondary education. Across the focus groups, teachers identified the challenges of measuring attainment and progression in skills, working scientifically and other outcomes not related to knowledge and understanding, and noted that these challenges were evident beyond schools.

> *I think a lot of the time it is based on the actual attainment of knowledge, and they mentioned skills loads – there are loads of Ofsted reports about how skills aren't being taught very well, but how are they judging that, how are they measuring it?*

There was broad consensus amongst teachers in the sample that measures of attainment should not be limited to knowledge and understanding.

Cognitive and creative

Perhaps unsurprisingly, the only cognitive and creative outcomes as defined by James and Brown (2007) found in the dataset were those relating to the creation of new personal knowledge by children, often linked to asking questions, solving problems, and designing and carrying out investigations, i.e. creating knowledge that was new to them. As one teacher put it:

> *They need to be confident and independent thinkers really and asking those big questions and knowing which route to go down to be able to answer those questions, "Do I need to answer this through research? Can I answer this through an investigation?"...for themselves getting to the point where they can make those choices and think independently.*

Another way this was expressed was in giving the children opportunities to plan their own investigations:

> *if we're giving them the experiment, it's not… we're taking some of the science away from them, whereas if we say, 'Right, we want to try and find out, you know, the answer to this question, how can we go about that?' So putting it to them as well for them to have a chance to design their own experiments.*

We did not find other examples of knowledge innovation (non-personal, cognitive or aesthetic, see table 6) being identified as priorities by teachers.

Dispositions

Two key dispositions were important to teachers who took part in focus group discussion: developing children's attitudes towards science and developing their curiosity about the world.

Attitudes toward science included interest in science, engagement with science, as well as enjoyment, relevance and inspiration. For example:

that's what you want them to have, a love of science and an interest in it because that's obviously what will take them further at the high school, rather than, you know, just thinking it's another subject.

It was important to teachers that this was not seen superficially as "a two minute wow at the beginning of the lesson" but about developing a more sustained interest in science. Some wanted to nurture an appreciation for science.

If anything, the last 18 months has highlighted to us why science is so key...I guess it's helping [children] recognise it's so intertwined in their lives, and that the opportunities it opens and enables, and how it can move our world forward for the better, and that each child can have an impact on that.

In terms of curiosity, teachers described the importance of science in encouraging children's questioning so that they want to find out more about science or the world. For example:

coming out of it with a sense of questioning why the way things are the way they are, and a willingness to investigate things further is… would be my sort of highest aspiration of what I'd be hoping to get out of it, but I know that that's not necessarily aligned with what we're necessarily supposed to see as the main priority.

Again, teachers identified challenges in connecting these dispositional outcomes to measures of attainment:

you can't label a child based on their ability to be curious and question and be an independent critical thinker...an assessed level, an attainment outcome at the end of a year might not actually appreciate the character that that child developed within themselves and their confidence to be able to go forward and plan things for themselves and make those mistakes..and develop that ability that, "We learn by these mistakes". We learn when it goes wrong, especially in Science…

Higher order thinking

In terms of higher order thinking, metacognition was discussed as an approach to achieve other outcomes (particularly attainment) rather than as an outcome in itself. The main outcome relating to higher order thinking was 'thinking like a scientist' related to scientific practices and the nature and culture of science. For example:

I've sort of tried to instil the idea of the importance of evidence-based research as a way of, you know, knowing what we know or basing our opinions on things that we have discovered through rigorous investigation rather than other sources, and I think it's similar to what has been said before about the importance of that in today's society in particular, over the last 18 months, the difference between like… you know, things that we know based on evidence and things that we know based on sort of spurious guesswork is quite important.

Outcomes in this category included awareness that science is based on evidence, that scientific knowledge changes depending on the nature and quality of evidence, that there are different ways of answering scientific questions, and how experiments work. This was often connected with both application and creating personal knowledge, as the following teacher described:

it's that ability to understand what it is they are investigating and to consider what will I need to use? What type of investigation will I do and what will I need to use to get a result? I

> *think for me they are the most important outcome, let them do it by applying things they already know so that they can really become scientists.*

Teachers discussed the difficulty in assessing outcomes related to thinking like a scientist, and knowing how that outcome has been achieved.

Membership, inclusion and self-worth

In the outcome category of membership, inclusion and self-worth, teachers described priorities relating to building science capital, ensuring that all children see themselves represented in science, and creating awareness of different types of careers in science (and in particular, local opportunities).

In terms of science capital, teachers discussed the importance of raising awareness of local industry and opportunities, bringing children into contact with people (including parents and STEM ambassadors) working in science, and making the science they learn personal to the children. For example:

> *One thing we've tried to focus a little bit more on over the last year and a half, two years or so is just recognising each child's science capital, and trying to incorporate that a little bit more into our planning and our teaching, and sort of personalising the science. Because I think the more we recognise that every child will access science in different ways and will have connections to science in different ways, the better.*

Teachers described scientists as often being seen as 'elitist' or requiring exceptional academic achievement, and saw their role as making science as a career accessible to children.

Representation was an important outcome related to membership, inclusion and self worth, particularly in relation to children seeing a future for themselves in science:

> *I think that's one of the big things, that if they've left primary schools and they don't see themselves represented in science, and they haven't had experiences where they can see themselves reflected in that, then they're not going to be that engaged in science in the future, or it'll be harder for secondary schools to change that attitude.*

Teachers described issues related to under-representation of women and black scientists, and the need to encourage children from all backgrounds that science is for them:

> *when we talk about, you know, the people working away to create vaccines for example, and we've tried to look at some of the individuals who've been involved in that as a way of trying to demonstrate the fact that it's all people from all different backgrounds.*

Understanding

Teachers valued outcomes relating to understanding science, including using scientific vocabulary, thinking critically and being able to relate school science to the world around them.

Across the focus groups, teachers described the importance of children being able to use scientific vocabulary accurately and appropriately in both spoken and written forms, including in relation to concepts (e.g. condensation, evaporation) and working scientifically (e.g. predicting, testing).

> *I think just get the scientific skills and a lot of the key vocabulary that they would be using in secondary. So basic, like knowing what prediction is, testing, results, and then obviously you know, the key facts that are in the National Curriculum.*

Teachers wanted children to be able to use vocabulary to describe what they had learnt and what they thought was difficult, and to make meaning of investigations.

> *It's also vocabulary, which I think underpins both skills and the knowledge. If they don't have that key scientific vocabulary, and they can't use it and have a discussion around it, then they'll have more challenges.*

Critical thinking was a valued outcome amongst teachers in the sample. This was taken to include asking deeper questions, making connections between different curriculum subjects, solving problems and being systematic.

> *So they are looking for patterns - researching, classifying, identifying, that kind of thing - so that they can use those skills in other areas of the curriculum as well.*

> *I think the main aim is critical thinking and to give them the big picture experience and I think it's how we develop units that don't constrain us like the National Curriculum constrains us.*

Outcomes relating to critical thinking were seen as desirable but difficult to assess.

A final aspect of understanding that teachers valued was about relating science to the world and (to a lesser extent) themselves. For example:

> *I think society as a whole needs to have an understanding of scientific literacy that enables them to engage in understanding what goes on, but also debate around things. You know, how can you talk about climate change and mobile phone technology and clinical trials if you don't have some sort of fundamental understanding of how things work.*

Using

Outcomes relating to using - i.e. the ability to develop scientific skills including practical skills and being able to apply them in appropriate contexts - were valued by teachers. Examples mentioned by teachers included researching, classifying, identifying and pattern-spotting.

Teachers discussed the use of scientific knowledge to ask questions, decide next steps in their learning and to improve their lives. For example:

> *I think it's one of the things where they learn about how to keep themselves safe, how to keep themselves healthy, how things work, and I just think that those skills are vital just like the Maths and English but I think it needs to be given more emphasis*

The teachers in the sample valued not only the development of practical skills, but children's ability to use knowledge and apply their skills in appropriate contexts, for example when deciding how a question can be investigated or designing investigations.

> *As you're developing the skill, you're building the knowledge at the same time, so that they can apply the knowledge to the skill that they're doing.*

This included applying ideas across topics, across other areas of the curriculum, and in secondary school.

Implications for the systematic review

Outcomes of importance for teachers beyond attainment and attitudes were identified in the focus groups. These are related to cognitive and creative outcomes, understanding and using science, higher order thinking (like a scientist), and membership, inclusion and self-worth. That said, there is considerable overlap between the keywords generated by teachers in describing their outcomes of importance, and those already identified in the protocol.

We therefore propose (i) that those keywords in Table 7 not already included as search terms are added to the terms, and (ii) that the outcomes listed in Table 7 are used in coding the outcomes of studies included in the systematic review. When reporting differential outcomes, for different groups of students, we propose including analysis by sex, race and ethnicity as well as by socio-economic background and for children with special educational needs and disabilities and those for whom English is an additional language (the latter three have already been identified for inclusion in the protocol).

Outcome	**Related keywords used by teachers**
Attainment	Knowledge, learning, memory, progression
Cognitive and creative	Inquiry, investigation
Dispositions	Attitudes, confidence, curiosity, engagement, enjoyment, excitement, interest, relevance
Higher order thinking	Evidence, thinking scientifically, working scientifically
Membership, inclusion and self-worth	Careers, diversity, representation, science capital
Understanding	Critical thinking, understanding, vocabulary
Using	Classification, identification, pattern-spotting, questioning, skills (and specifically 'practical skills')

Table 7: Keywords related to outcomes valued by teachers

Approaches used in primary science education

The approaches taken by teachers to curriculum, teaching and assessment are summarised below. There was variation in the amount of time teachers had allocated to science - but this tended to be less than the time allocated to English and Mathematics, typically between one hour and half a day per week.

Curriculum

Teachers in the focus groups described science being planned into the curriculum in four main ways.

As part of thematic units or topics. This was where science was integrated with other subjects such as history, in order to examine a big idea such as 'flight' from different perspectives:

> *We've been redeveloping our whole curriculum recently the past couple of years and we're trying to make it very interactive for the children. It's based on a dramatic approach so that...this term for example, our overarching focus is history...in our year 1 we're talking about flight – so let's say I am the Wright brothers' mother for example. So you are kind of trying bring into for them the topic alive so that they can kind of remember and learn. So through that we're doing Science that's, we're doing materials but we're also trying to link it into flight.*

Driven by scientific knowledge. This was when teachers described prioritising knowledge acquisition in their curriculum design.

> *We have got as a school quite an emphasis on knowledge...in terms of what children learn by the end point, our curriculum is split down in small steps that we would like the children to know, so for a unit that has perhaps got four national curriculum objectives, there would perhaps be 20 pieces of component knowledge which make up the composite which is those four national curriculum objectives.*

Driven by 'working scientifically'. This was when teachers described prioritising scientific skills and disciplinary practices in their curriculum design. In the example below, this facilitated working in classes of mixed year groups.

> *I start with working scientifically skills first and they are a priority and I have developed what we call our master rewards, and I break them down...so there are six areas of working scientifically….I want scientists who are proficient in working scientifically. Their knowledge it is hard to define it by the end of Year 6, because they do things on alternate years. They will have achieved a certain amount by the end obviously at that point, but that is just what is covered in the national curriculum. My key objective is the working scientifically skills.*

Integrating knowledge and working scientifically. This was when teachers described prioritising both knowledge acquisition and working scientifically, and making a concerted effort to ensure both are developed every lesson.

> *We're trying to shift towards making sure the children have that substantive knowledge before they then carry out investigations because I think in the past there has been a danger of activities being done without the learning being there and the children don't remember…. so, we're really concentrating now on making sure that staff are doing it that way around, if you like, rather than starting with a hook. These things – I just think sometimes, "Oh, it's really exciting; we'll do this to kick it off", but the children don't know what they're doing so it's lost really.*

The choice of curriculum design is likely to be linked to the outcomes being prioritised by the school. Whilst it is unlikely that there will be clear comparisons between different approaches to curriculum, a potentially useful thread for the systematic review to follow is the link between curriculum approach and pupil outcomes.

Assessment

Assessment was an area of primary science practice where there was some appetite for evidence. Even where exemplars were used in standardisation and moderation was perceived to be robust, there was some dissatisfaction or lack of confidence in approaches to assessment (reportedly common amongst wider networks of science subject leaders). As one teacher put it:

> *if there is some actual research-base, this has worked really well to give you a good assessment toolkit, then I think teachers would love to have that information available.*

There was a wide variety of approaches used in schools, including self-assessment, peer-assessment and teacher assessment, and there was concern to not over-burden teachers with assessment, particularly when so little time was given to teaching primary science. There was some concern amongst teachers that there was a lack of clarity of what 'greater depth' looks like in science, with the result that judgements from English and mathematics were being carried over into science for reporting science outcomes.

Self-assessment and peer-assessment often took the form of children highlighting what they can do in their books. For example:

> *We have them all in the front of the children's books so they can also see what they are doing when they're working scientifically. As we go through each topic area we highlight in yellow initially, and then we go over them in green with a green highlighter if they've done them more than once, so obviously they'll hit them two or three times doing the different topics throughout the year. And the children can assess them themselves.*

Teacher assessments included the use of concept maps, concept cartoons, discussion tasks, writing tasks, low stakes quizzes, observations during lessons (supported by a teaching assistant), the Teacher Assessment in Primary Science (TAPS) resources, and other commercial packages. for example:

> *we set up questions and ask children to explain their reasoning, a bit like concept cartoons, "This character says this; do you agree with it? Discuss why or why not", and that sort of thing*

> *we used the TAPS programme, so the Teacher Assessment in Primary Science, and so every half term all children all complete a focussed assessment, a practical one, because it was that working scientifically element that we found much more tricky to evidence. And that's recorded. It depends though; in reception it'll be more around a floor book and they will do something around that; that would be the sort of evidence we use there. Our Teachers will also record videos and use a QR code so that they've got that to refer back to to capture those rich discussions….the evidence can look very different depending on what the focus is but there would be something working scientifically at least once a term.*

Teachers in the sample tended to prioritise assessment that was used formatively as part of feedback to children, and to avoid burdensome processes linked to reporting.

Teaching

Approaches to teaching were coded using the recommendations from the Education Endowment Foundation's Improving Secondary Science guidance report as a coding framework. We found teachers reporting approaches based on children's preconceptions, self-regulation approaches, the use of practical work, and references to the importance of memory and language of science. Feedback was not discussed in relation to teaching approaches, and no references to modelling were found in the focus group transcripts: this does not mean that feedback and modelling are not used, only that they were not mentioned as teaching approaches. In the limited time available, it was not possible to cover all teaching approaches, and the examples teachers tended to draw on were recent examples.

- Language of Science: Develop scientific vocabulary and support pupils to read and write about science.

Approaches to developing children's use of scientific vocabulary were widely used amongst the teachers who participated in the focus groups. This included sharing key vocabulary, providing opportunities for children to use vocabulary in oral explanations and in writing, using knowledge organisers and (for younger children) floor books, and the approach described below:

We do is something called, 'Words I've Never Heard,' which is what we do in reading, and in our writing unit with our model of excellence, so what we do is we do that for the vocabulary for science, so we'll choose six words so it might be condensation or evaporation, and we look them up and we talk about them and we discuss where we've found them before and we see if we can use them in the sentence, we look at the spellings of them because spelling's an issue in this school...a lot of the scientific words are very challenging for our children, so we spend a lot of time on that.

This was a particular challenge for some children, for example:

We have a lot of language issues. So, our white British children are the ones that actually have almost little English skills, not our EAL, and it's trying to do a lot of Science talk in early years. And a lot of our recovery curriculum went on...just Science talk. So, everything was more speaking and listening really because they wouldn't have done any sort of talk at home during periods of lockdown let alone it being Science.

- Memory: Support pupils to retain and retrieve knowledge.

Teachers described approaches to teaching science which encouraged retrieval of knowledge from previous units. Examples included asking children to use sentence stems to make links between one unit and another, and classroom talk about relevant prior learning experiences. These were often starter activities.

We have a 'Do Now!' which would be a knowledge retrieval from either another unit or some other, or something from another year group,

We do something called 'Keeping Skills' in our school where it's like a starter activity. It keeps substantive knowledge ticking over from previous years...you might have something on forces or plants or whatever and it's revisiting the previous curriculum or maybe even two years ago so that knowledge is ticking over so that then connections are made when new topics are launched.

- Practical Work: Use practical work purposefully and as part of a learning sequence.

Practical work was widely referenced as an approach to teaching primary science, along with concerns about availability of time, resources and laboratory space which limited the types of practical work that could be carried out (investigative practical work was described as taking longer). There was concern that practical work should be purposeful, and linked to learning, for example:

So they have a real excitement around experimenting and finding things out, but it needs to be, that's the difficult part of it, sort of honing it in to actually what are you, what have you been looking for, what are you, you know, what have you actually learned?

There was also concern to ensure that children had opportunities to think about what they were doing and not just follow instructions from the teacher, often in the context of enabling children to design and carry out their own investigations:

Child-initiated experiments because a few years ago the teachers were telling the children what they were doing: "this is what we're investigating and these are the resources you will use and this is what you will do". But then I said to Teachers, "What about those children that don't want to investigate what you want to? We need to try and incorporate some of the principles from early years Science into our teaching where the children think of their own questions and they get their own resources and then they can present their findings to an audience based on their investigation", which obviously promotes independent learning, it *increases the engagement and enjoyment of children as well; they're pursuing their own interests rather than following the teacher's lead so they're more independent.*

- Preconceptions: Build on the ideas that pupils bring to lessons.

Preconceptions were not mentioned directly by participants, but approaches to identifying misconceptions such as the use of multiple choice questions and concept cartoons were discussed.

> *We can use the various bits of software to identify common misconceptions and things like that, we take those and put them into kind of teacher training sessions so that what we think they've struggled with in the last round of assessments will be brought up, and then then embedded into upcoming lessons even if it's on a different unit, it's kind of you know, recapping prior learning but as a way of addressing misconceptions.*

There was concern from some teachers about their own misconceptions which had been spotted by secondary colleagues.

- Self-regulation: Help pupils direct their own learning.

In this category, teachers mainly described metacognitive approaches which they had been introduced to via the Education Endowment Foundation. They tended to have experienced non-science specific professional development, which they then applied to science.

> *When you speak to our children it is those metacognitive approaches that when you are talking about attainment and enjoyment – because to me if they are enjoying it and you have got the behaviours right and you have got everything around the metacognitive approach correct, the science, the attainment within science and the enjoyment within science is for me as we have seen that progress over the last few years...it has worked wonders for some of our children, and especially those children who you would class as your most vulnerable*

In addition to the approaches from the improving secondary science guidance report, teachers reported the following approaches:

- Involving children in school-level science opportunities, including children from different years participating in science fairs and training science ambassadors in Key Stage 2 to work across the school.
- Working with visitors, including scientists in the local community and amongst school family members as well as STEM ambassadors. The nature of interaction between visiting speakers and children was not always shared, but there was considered to be value in making these connections.
- Approaches linked to outdoor learning, including gardening, outdoor investigations and forest schools.

> *We have done a lot of work as well in forest school. We are trying to when we are doing our investigating get outside as much as we can, because we found that if we do something outside and really hands-on and practical, the children are retaining the information a lot better. So, for instance we did things like when we are doing seed dispersal, we have actually gone outside with balloons and seeds and we have popped the balloons, and let the seeds fall on the ground then we have observed over time and wanted to see if these seeds have germinated.*

Influences on primary science education

'Influences' includes reported barriers and enablers to primary science education, responses to the educational impact of the COVID-19 pandemic, leadership and professional development in primary science, and use of research evidence. Barriers and enablers are reported by theme because depending on school context, an influence could serve as either a barrier or an enabler. Nine influences are presented in Table 8.

Influence	**Teachers' perspectives on how the influence functions as a barrier or enabler**	**Example quote(s)**
Assessment policy	● Absence of end of Key Stage 2 Science assessment liberates teachers to go beyond the curriculum. ● Non-assessed subjects (including science) are squeezed, particularly in year 6. This limits time available for investigative work.	*When you look at science sampling tests that are done every other year by the government...there is huge disparity [between sampling tests and teacher assessment] and ...what is the expected level.*
Black Lives Matter	● Encouraged teachers to think about the diversity of scientists children see in class. ● Websites now available to support science teaching.	*With Black Lives Matter, we've really thought about diversity and inspirational people…each teacher has to choose three inspirational people that covers the diversity as well.*
Evidence	● Metacognitive approaches and those targeting memory were reported to support primary science education. ● Evidence on primary science education was not considered accessible. Teachers who discussed using evidence drew on the EEF Improving Secondary Science report, Rosenshine's Principles of Instruction, academic study (Masters), and other evidence digests including blogs.	*It is those metacognitive approaches...if you have got everything around the metacognitive approach correct, the science, the attainment within science and the enjoyment within science...it has worked wonders for some of our children, and especially those children who you would class as your most vulnerable as well.*
Experiences of COVID-19	● Disruption of plans: unable to host visitors and science fairs; allocation of children to 'bubbles' meant delays to introduction of a science ambassador programme. ● Teaching assistants deployed to prioritise 'catch-up'. ● Cleaning protocols made it difficult to share resources and physical distancing reduced practical work. ● Greater appreciation for science and the impact that scientists have on everyday life. ● In some cases, more contact between school and home. ● Increased use of online technologies for professional development, networking, linking with scientists and collaborating with partner schools. ● Smaller groups in school and use of outdoor space provided new opportunities to do science. ● Virtual science club allowed more children to participate.	*The funding implications mean that the TAs across the school..are being used to do interventions because of Covid catchup or supporting children who have specific needs. There are quite a few children that we have here who have either behaviour issues or learning issues and we don't have money for them.* *One thing I will keep is the zoom meeting with my partnership schools because that has been a revelation. In the sense that we're all spread out all over the county and it was a nightmare trying to get everyone to drive to one place… I'll definitely keep that. And I have enjoyed all these zoom things because you do get to chat to lots of different people that you would never ever normally meet.*

Influence	Teachers' perspectives on how the influence functions as a barrier or enabler	Example quotes
Leadership	Leadership acted as an enabler where they: ● Allocated curriculum time to Science (one hour/week typical; some schools allocated a morning or afternoon each week). ● Enabled science leads from across trusts to work together (but trusts could also be considered protective of their approaches). ● Included science in the school development plan. ● Required science leaders to report to governors. ● Resourced subject leaders with time for their role (including for professional development and monitoring).	*We have got quite a good set up in the sense we've got an Assistant Head in charge of curriculum. She's very supportive and...as a result, things do get timetabled into staff meetings. I get released from my lessons to do stuff so at the moment I'm trying to get PSQM action plans.. I think that's really important...otherwise you're doing it all at home in your own time and it's too much.*
National curriculum	● Constrains what can be taught (described as narrow, prescriptive, outdated, with lack of challenge particularly at Key Stage 1). ● Ensures children are prepared similarly for secondary school (although concerns about how primary science is recognised at secondary level). ● Framed around content that does not always link to the 'real world'.	*The National Curriculum is a minimum expectation and actually everybody is working towards it being this is what you should do to deliver the National Curriculum, and actually what we're all saying is it's a bit rubbish. Can we just have a better National Curriculum?*
Nature of the cohort	The nature of the cohort of children being taught was also identified as important to teachers. For example: ● Prior attainment was used to inform planning ● Parental engagement and resources had enabled children to do (or not) 'kitchen sink science' during lockdowns. ● Teaching was described as more difficult when children were unfamiliar with vocabulary e.g. plants or animals or prior experiences e.g. beach visits. ● Meeting the needs of all children in the class.	*We have got quite a high percentage of SEN in our school..so, it's including those children in our investigations..quite often with these children they'll be so excited by the investigation it's how to actually get these children to take home a message from the lesson and not, "We've mixed things together and we've had great fun with it". ...we have to pick what our priority is for these children because we very much want them to be included in the process.*
Ofsted	● Those in schools judged good and outstanding felt they had freedom to innovate. ● Perceived to confer lower status on science than English and Maths. ● Workload associated with changing Ofsted priorities, mock inspections, deep dives and interviews.	*I'm talking about Ofsted being the guide to our living existence. It used to about about data and it used to be about interviewing the headteacher. Now it's all gone back to watching teachers again.*

Table 8: Influences on primary science education (continued)

Influence	Teachers’ perspectives on how the influence functions as a barrier or enabler	Example quotes
Professional development	• Primary teachers often do not have a science background and might need support with confidence, subject knowledge and pedagogy. • Little time on Science in initial teacher education. • National organisations*[1] provide funding for science and professional development. • Professional development (specifically in subject leadership) can be expensive. • Conferences, partnerships, regional networks and Twitter provide support. • Secondary teachers challenge misconceptions of primary teachers and provide resources, advice on practical work and access to labs. • Visitors (e.g. STEM ambassadors, Planetarium) provide new experiences for children. • More freely available online live lessons for children and CPD for teachers is now available.	*You get so much CPD through PSQM and so many opportunities to network with other science leaders, I feel like I have had a lot. But in terms of whole teaching staff CPD over the last 18 months, other than through me...not external for the last two years.* *Pre-pandemic I think it was really rare that you would do this kind of thing [network with each other]. Now, it seems like there is so much professional development out there..the opportunities that that then provides to the children are far greater than what they were...It is so easy to get out there now, and take your school out there to other schools and bring other schools to you, share ideas, improve curriculum and improve pedagogy.*
School resourcing	• School funding formula means that schools have different opportunities to resource primary science. • Size of school gives some schools economies of scale. • Pupil Premium funding supports a lot of work - but can change the nature of the intake and limit later opportunities. • Restricted time for science can limit practical work. • Tight budgets make it difficult for schools to buy quality schemes of work, resource teachers with membership of subject associations and professional development, and purchase sufficient resources and consumables for all children. • Lack of laboratory infrastructure.	*I think we’re lucky at our school because we are a big school. It means that all the subjects are spread out between the staff. Whereas I have colleagues who work in small schools and they’re doing more than one role. So they might have Science, but they might also have another subject as well. And that’s, I think that must be very difficult cos although you’ve only got a smaller number of children, you’ve still got as much development to do with your subjects. So I think that’s where we benefit as well.*

Table 8: Influences on primary science education (continued)

[1] Specific organisations named included: the Association for Science Education, Education Endowment Foundation, Natural History Museum (Explore: Urban Nature), Ogden Trust, Primary Science Quality Mark (PSQM), Primary Science Teaching Trust (PSTT), particularly Teacher Assessment in Primary Science (TAPS) materials, and Wellcome Trust (Explorify).

Recommendations for the systematic review

Based on the research on practice, we recommend:

- Inclusion of an expanded range of outcome types for inclusion in the systematic review (connected to attainment and attitudes) where the approach takes place in a primary science context and/or the outcome measure relates to primary science.
- Inclusion of the following keywords in the searches being conducted for the systematic review.

Outcome	**Related keywords used by teachers**
Attainment	Knowledge, learning, memory, progression
Cognitive and creative	Inquiry/Enquiry, investigation
Dispositions	Attitudes, confidence, curiosity, engagement, enjoyment, excitement, interest, relevance
Higher order thinking	Evidence, thinking scientifically, working scientifically
Membership, inclusion and self-worth	Careers, diversity, representation, science capital
Understanding	Critical thinking, understanding, vocabulary
Using	Classification, identification, pattern-spotting, questioning, skills (and specifically 'practical skills')

- Coding outcomes of studies included in the systematic review using the outcome types identified in the table above.
- Where 'curriculum' is included in articles included in the review, it will be useful to classify as close as possible to the four models teachers described: cross-curricular themes, knowledge-led, working scientifically-led, or integrating knowledge and working scientifically.
- The review should include studies which investigate effective approaches to assessment in primary science, linked to the outcomes above, and analysed with reference to sex, race and ethnicity, EAL, special educational needs and socio-economic background. Disadvantage will not be a requirement for inclusion, because too few studies would remain. It will be used in the analysis and synthesis.
- The report should include commentary on the applicability of recommendations from the Improving Secondary Science guidance report to primary science education, based on the studies included in the systematic review of primary science.
- The influences on primary science should be used as contextual information in the report of the systematic review and preparing the guidance report, so that the research findings are interpreted for, and sensitive to, the English primary science education context.

References

Brundrett, M., & Rhodes, C. (2013). Theories of Educational Research. In M. Brundrett & C. Rhodes (Eds.), *Research educational leadership and management* (pp. 12–23). London: Sage.

Education Endowment Foundation. (2018). *Improving Secondary Science. Guidance Report.* Education Endowment Foundation. https://files.eric.ed.gov/fulltext/ED612065.pdf

Grant, M. J., & Booth, A. (2009). A typology of reviews: an analysis of 14 review types and associated methodologies. *Health Information and Libraries Journal, 26*(2), 91-108. https://doi.org/10.1111/j.1471-1842.2009.00848.x

James, M., & Brown, S. (2007). Grasping the TLRP nettle: preliminary analysis and some enduring issues surrounding the improvement of learning outcomes. *The Curriculum Journal, 16*(1), 7-30. https://www.tandfonline.com/doi/full/10.1080/0958517042000336782

Appendix 3: Theoretical review

Effective primary science teaching: a theoretical review

This theoretical review of primary science teaching is intended to feed into and to complement our systematic review of primary science teaching. Its particular foci are to:

- Explain how children learn primary science;
- Discuss the mechanisms by which models of learning in primary science are presumed to work.

'Effective' is understood with regards to the acquisition of: knowledge; understanding; skills; and positive attitudes towards science / school science. However, the literature does not distinguish consistently between these, and very few studies consider all of them. It is also the case that much literature does not distinguish between pupils, except in terms of gender, so that it is difficult to say a great deal about disadvantaged learners that goes beyond gender.

This is not a systematic review, though it proceeded in parallel with our systematic review. References were obtained from a range of sources:

- Existing knowledge of the primary science education literature within the team, including well-regarded sources of advice to teachers of primary science;
- Additional literature obtained as a result of searches resulting from our existing knowledge, our review of practice, emerging findings from our systematic review, and suggestions from others.

What is science?

Some discussion as to what is meant by 'science' is needed before deciding what makes for effective science teaching. There is a very large literature on what science is in both the history and philosophy of science literature and the science education literature.

The word 'science' comes from the Latin scientia ('knowledge') but there are many non-scientific kinds of knowledge. Science is generally understood to be about knowledge gained by empirical studies of the material world. All cultures, from their earliest recorded history, practised science, and many Western scientific 'discoveries' were in fact known about elsewhere long before (Reiss, 1993).

Merton (1973) characterised science as open-minded, universalist, disinterested and communal. This means that scientific knowledge should be objective – that is, it should not matter who is producing it. Of course, that is an oversimplification; different scientists have different interests, and the area of focus for scientists is generally determined by who is paying for the work. Nevertheless, scientists should always be prepared to change their views in the light of new evidence, better interpretation of existing evidence or the production of new, more fruitful explanatory theories.

Certain things clearly fall under the domain of science – the nature of electricity, the arrangement of atoms into molecules and human physiology, to give three examples. However, what about the origin of the universe, the behaviour of people in society, decisions about whether we should build nuclear power plants or go for wind power, the appreciation of music and the nature of love, for example? Do these fall under the domain of science? Some people would argue 'yes' to all of these questions and the term 'scientism' is used, pejoratively, to refer to the view that science can provide sufficient explanations for everything. However, most people hold that science is but one form of knowledge and that other forms of knowledge are needed to complement scientific knowledge (Reiss, in press).

While historians tell us that what scientists study changes over time, there are reasonable consistencies:

- Science is concerned with the material world, whether natural or manufactured. So, the laws of gravity apply as much to artificial satellites as they do to planets and stars and apples.
- Science is concerned with how things are rather than with how they should be. So, there is a science of chemistry and endocrinology without science telling us whether chemical warfare or the use of human growth hormone are good or bad.

The phrase 'nature of science' (NOS) is used in science education to describe the essence of science – what science is all about. NOS is therefore about the epistemology of science, science as a way of knowing. At its crudest, we can think of science being both about a way of generating knowledge (the 'methods' of science) and as a body of knowledge (the 'content' of science). These two aspects, sometimes referred to as 'disciplinary' knowledge and 'substantive' knowledge, respectively, are reflected in many science curricula around the world, existing in the latest version of the science National Curriculum in England and Wales as 'Scientific knowledge and conceptual understanding' and 'The nature, processes and methods of science' (Department for Education, 2015).

In line with school science curricula, we take science to mean 'the natural sciences', so that it includes astronomy, biology, chemistry, earth science and physics and excludes the social sciences, except that some psychology falls within the natural sciences. However, psychology does not feature in primary science curricula.

Pre-primary science

Primary science can straightforwardly be taken as the science that a child learns while in primary / elementary school. In England, the age range is therefore from about 5 to 11 years, though the age at which children start, and finish, their primary education varies from country to country. It is also, of course, the case that during their primary schooling children learn much of their science from non-school sources (e.g., Tunnicliffe & Reiss, 1999, 2000; Riedinger *et al.*, 2011).

Although the time before children arrive at primary school is outside the scope of our systematic review, children learn about the material world from the time they are born, and their pre-birth experiences are relevant too. The term 'emergent science' is often used to characterise children's learning of science from birth up to a less precisely defined age (which varies from about three years through to about eight, depending on author). There is an open access Journal of Emergent Science (https://www.ase.org.uk/resources/journal-of-emergent-science) that began in 2011 and there are a large number of books (e.g., Johnston, 2005, 2014; Russell & McGuigan, 2016) and articles that examine it. Writings on pre-primary science often give a particular emphasis to play, to which we return below in a section of its own.

In addition to literature on pre-primary science education, there is a long tradition within developmental psychology of studying how pre-school children's learning develops. Unsurprisingly, much of this is about how young children learn to speak, to develop motor coordination and to relate to others but there is a body of literature about young children's understanding of the wider world. Much writing by child psychologists emphasises the tremendous capacity of young children to learn: "Babies' brains seem to have special qualities that make them especially well suited for imagination and learning" (Gopnik, 2009, p. 11).

Even children within a year of birth can be shown to 'know' certain things about the world. Given that children of this age typically cannot yet speak more than the occasional few words, much of the evidence for what they know relies on the useful observation that babies and infants spend longer looking at the unexpected. Babies as young as three and a half months spend substantially less time staring at an experimental set up that shows a moving short carrot that disappears behind a short wall and then re-appears than they spend staring at an experimental set up that shows a moving tall carrot disappearing behind a short wall and then re-appearing (Goswami, 2008). In everyday language they 'know' that part of the tall carrot should have been visible above the short wall.

Using this and other techniques, psychologists have established that pre-school children know the following:

- Objects exist continuously in time and in space (so, for example, a rabbit cannot suddenly appear, disappear or change greatly in size or move instantaneously from one place to another).
- Objects need to be supported to avoid falling downwards (so a box falls if the table on which it is standing is withdrawn).
- Changes to inanimate objects are generally the result of causes whereas animals are agents who can move on their own (so it is not surprising when a dog moves without anything coming into contact with it but it is when a meaningless shape does).
- Causes precede effects (so if a marble is to cause a jack-in-the-box to appear it needs to be dropped into the apparatus before the jack-in-the-box appears).
- If an effect has a number of potential causes, the actual cause is likely to be one that covaries with the effect (so by the time they are three years old, children can work out which of a pair of levers on a box causes the light on the lid of the box to come on).
- Causes and effects must be contiguous in time and place.
- By and large, causes and effects are of similar types (so a mechanical effect, such as a change in movement of an object, is more likely to be caused by a mechanical cause, such as its collision with another object, than by a flash of light or the appearance of a smell).
- The behaviour of objects is frequently the result of intentions (so in a computer display that initially shows (i) a large circle and a small circle separated by a tall rectangle, then shows (ii) each circle in turn expanding and contracting twice and then shows (iii) the small circle moving towards the large circle, reaching the tall rectangle, retreating, moving again towards the large circle but this time passing above the rectangle and making contact with the large circle is interpreted by adults and infants alike as a mother (large circle) calling to her child (small circle) who then runs towards her, only to be prevented by a barrier, which causes the child to retreat so as to be able to run towards the barrier and jump over it and thus reach mother, whereupon mother and child embrace). Even three-month old babies are sensitive to movement information that is interpreted by adults as specifying social causality.
- Artefacts do not grow but organisms do (so small kettles do not increase in size over time whereas rabbits do).
- Offspring have characteristics that resemble those of their parents (so when four-year-olds are shown a picture of a newborn kangaroo – a shapeless blob – and then told that it was raised with goats they are almost all sure that it grows up to be good at hopping not climbing).

(Abrahams & Reiss, 2012a, p. 414)

We now move to examine what the literature says about effective primary science teaching. To emphasise, we do not here review the evidence as to how effective various approaches to primary science teaching are – that falls within the remit of our systematic review. Rather, what we do is examine the various approaches that are described in the literature as being effective, paying particular attention to why each approach might be effective, i.e., to the mechanisms that are presumed to underlie it.

We list the various approaches very roughly in the frequency with which they are advocated in books, reviews and reports about primary science teaching, from the more frequent to the less frequent.

Practical work

The term 'practical work' is used in science education to refer to any type of teaching and learning activity in which learners, working either individually or in small groups, are involved in manipulating and/or observing real objects and materials (e.g., determining which of a selection of objects are magnetic, or observing rocks under magnifying lenses to classify them as igneous, sedimentary or metamorphic) as opposed to virtual objects and materials such as those obtained from a DVD, a computer simulation or a text-based account (Millar, 2011).

Practical work is often seen as a key feature of science. Although some countries place a lower emphasis on the importance of practical work in school science (e.g., Germany – di Fuccia *et al.*, 2012), in a number of countries, secondary schools have dedicated rooms – school laboratories – for the teaching of science; even when general classrooms are used for the teaching of science, as in most primary schools, there is specialised equipment to enable practical work to be undertaken. Practical work thus helps distinguish school science from other subjects, though some other subjects also have specialised equipment (e.g., art, music, physical education) and sometimes take place in dedicated spaces.

The importance accorded to practical work in school science, for learners of all ages, is not surprising, given that science, as discussed above, is both about a body of knowledge about the material world and a set of methods (sometimes referred to as 'practices') that are used to derive this knowledge. Practical work, both in schools and as practised by professional scientists, therefore entails using specialist equipment to interrogate aspects of the material world.

A wide range of aims for practical work have been proposed. Bennett has suggested that despite a certain degree of variation between studies, there is a general consensus, whatever the phase of education, among teachers that the most important aims of practical work are:

- to encourage accurate observation and description;
- to make scientific phenomena more real;
- to enhance understanding of scientific ideas;
- to arouse and maintain interest (particularly in younger pupils);
- to promote a scientific method of thought.

(Bennett, 2003, pp. 78-79)

These aims overlap considerably with those in an earlier list, proposed by Hodson:

- to motivate, by stimulating interest and enjoyment;
- to teach laboratory skills;
- to enhance the learning of scientific knowledge;

- to give insight into scientific method, and develop expertise in using it;
- to develop certain 'scientific attitudes', such as open-mindedness, objectivity and willingness to suspend judgement.

(Hodson, 1990, pp. 33-30)

Rather less attention has been given to the mechanisms through which these various aims of practical work in school science might operate. A somewhat cynical response to the aim of practical work to motivate learners might be that in most subjects, school pupils spend their time either listening to their teachers or writing. Accordingly, practical work provides a welcome diversion. The word 'motivate' is often used somewhat loosely, rarely being distinguished from engagement or interest. Abrahams concluded from a study undertaken in secondary schools in England that "whilst practical work generates short-term engagement, it is relatively ineffective in generating motivation to study science post compulsion or longer-term personal interest in the subject, although it is often claimed to do so" (Abrahams, 2009, p. 2335).

In a controlled trial to evaluate the effects of the 'Conceptual Challenge in Primary Science' project, Mant *et al.* (2007) found that science lessons developed by teachers, after continuing professional development (CPD), that had more practical work, more discussion, more thinking and less (but more focused) writing, resulted in both pupils and teachers reporting greater engagement and motivation and to increases in pupil attainment.

In a study to investigate possible benefits of peer tutoring for primary science, Year 5 pupils, under the supervision of pre-service teachers, collaborated with Year 2 pupils and worked their way through plans supported by the resources they had prepared (Stephenson & Warwick, 2001). In addition to improvements to investigative skills, there were also clear benefits to other, more generic skills, notably communication, working with others and improving their own learning and performance.

Much of the research on putative benefits of practical work has looked at the consequences of practical work for conceptual knowledge. It is now widely accepted that in both primary and secondary schools, students can engage in practical work in science and enjoy it but gain almost nothing from it in terms of conceptual learning. In their review of primary science in England, Bianchi *et al.* (2021) conclude that "Children retell the 'magic' moments in science learning and aren't able to explain what they have seen or the concept explored" (p. 6). A distinction is now often made between the 'hands-on' doing of science (undertaking the practical activities) and the 'minds-on' learning (of conceptual knowledge) that may or may not result, a distinction that builds on the work of Tiberghien (2000).

Using a framework derived from Tiberghien (2000), Abrahams and Reiss (2012b) found that primary science teachers spent a substantially higher percentage of their science lessons talking with pupils about the scientific ideas associated with practical activities than did secondary science teachers, whereas students in the secondary lessons spent a substantially higher percentage of their science lessons engaged in manipulating objects and materials. Abrahams and Reiss suggested that practical work might be made more effective, in terms of developing learners' conceptual understanding, if teachers adopted a more hands-on and minds-on approach and explicitly planned how learners were to link these two essential components of practical work.

In a follow-up study, Abrahams *et al.* (2014) evaluated the work of the 'Getting Practical: Improving Practical Work in Science' CPD programme that was undertaken by the Association for Science Education in both primary and secondary schools in the UK and was indeed intended to ensure that practical work in school science was minds-on as well as hands-on. What they found was that whilst the CPD programme was effective in getting teachers to reflect on the ideas associated with the Getting Practical programme, it was much less effective in bringing about changes in actual teaching practice. It was notable that in the one school (of 30: ten primary and 20 secondary) in which there was substantial impact, the teacher was a well-established and respected Head of Department, one of only two in the sample:

> As a consequence of her seniority she was able to drive the Getting Practical CPD ideas forward within the department and to do so over a sustained period of time in a manner that it has been suggested (Joyce & Showers 2002; Loucks-Horsley *et al.* 2010) is necessary if lasting change to teaching practice is to occur, rather than seeking only to raise awareness amongst teachers of new ideas. Indeed, Loucks-Horsley *et al.* (2010) suggest that sustained support would 'include 30–100 contact hours over a time period ranging from 6 to 12 months' (p. 123), both of which were achieved by [this teacher]. One example of such sustained support was her introduction, into the science department's teaching timetable, of regular peer observations of practical work to facilitate intra-departmental reflection on new effective practice. Certainly, by the time of the post-CPD interview these peer observations had been going on regularly for almost a year.
>
> (Abrahams *et al.*, 2014, p. 276)

Other studies have found that it is difficult to change science teachers' practical work practices. Correia and Freire (2016) evaluated a one-year professional development programme which aimed to promote the use of practical activities in classrooms by elementary teachers. They found that most teachers were able to overcome their initial difficulties and progressively gained more confidence. However, one year after the end of the programme, teachers reported that their actual practices had not changed significantly.

Inquiry-based science education

Related to practical work is inquiry-based science education (IBSE) but here there is an explicit commitment to an investigation that is open-ended, unlike many practical activities in school science which are sometimes characterised as 'recipe following'. Harlen and Qualter give an example as to how a science inquiry for primary pupils might be arrived at and proceed:

> It started with a question, stimulated by a story, and made real through a collection of different balls. The initial question 'which ball is best for a dog?' was turned into an investigable question: 'which is the bounciest ball?' The children made lots of suggestions based on their previous experience of balls ('The red one. It looks like it's bouncy') and they suggested explanations ('The rubber might make it bouncy'). Their various predictions as to which might be the best ball were then challenged by the teacher's question: 'How do you know?' The teacher asked for the children's ideas on how to collect data that could provide evidence of 'bounciness'. Through discussion of many different suggestions they agreed the procedure, which was in two parts. In the first part, they found four balls which they judged to be the bounciest. The second part comprised a further round of planning and data gathering with these four balls. Then, the final result was compared with their predictions. The record of the whole inquiry enabled the children to look back and reflect on what they had done and learned, not just about balls but how to answer a question through scientific investigation.
>
> (Harlen & Qualter, 2018, p. 135)

At its inception, the science National Curriculum for England and Wales had a commitment to scientific investigation. One of the original 17 Attainment Targets of the inaugural (1989) science National Curriculum (AT1) was titled 'Exploration of Science' and in the 1991 simplification from 17 Attainment Targets to four, it survived, combined with the old AT17, 'The Nature of Science', as 'Scientific Investigation' (Sc1). This commitment to scientific investigation arose from a long history in the UK of inquiry in school science, fostered by the Nuffield Foundation and others. As Jenkins puts it:

> The incorporation within a national curriculum of an Attainment Target concerned with scientific investigation can be regarded, therefore, as a statutory codification of a longstanding curriculum commitment to science as an activity concerned with 'finding out' or discovery, based on experimentation as a means of generating new knowledge and understanding of some aspect of the natural world.
>
> (Jenkins, 1995, p. 472)

As Jenkins goes on to say, given this longstanding curriculum commitment, "science teachers in England and Wales might be expected to have welcomed scientific investigation, if not necessarily its interpretation as Sc1, as reflective of a fundamental and established feature of their professional practice" (p. 472). However, Sc1 proved difficult to implement and assess. As one teacher put it:

> ... if this whole thing had been our idea, it would have been different. None of it was our idea, not any of it, it was all government decision and they appointed people to do this. What they have done is ... come up with paperwork with fantastic ambitious ideas and they have not managed to put any of it into schools in a usable form. They can't, because we are not on the same wavelength.
>
> (Jenkins, 1995, p. 476)

It is fair to say that subsequent revisions to the National Curriculum have attenuated to the point of eliminating the requirement for science inquiry. The cause of science inquiry in both primary and secondary schools has not been helped by considerable controversy as to its efficacy. As sometimes happens in education – think the controversies over 'New Math' (Reys, 2001) and phonics (Torgerson *et al.*, 2019) – 'inquiry' has become something of a touchstone both in science education and more generally. While many primary science educators remain committed to it – as the above quotation from perhaps the most respected textbook on primary science (now in its seventh edition) indicates – more politically conservative commentators see it as indicative of an abandonment of proper instruction, replaced by a commitment to a child-centred form of teaching. Christodoulou (2014) identifies 'Projects and activities are the best way to learn' as one of her seven 'Myths about education'.

Even those who advocate IBSE (though not to the exclusion of other ways of teaching science) acknowledge the difficulties in doing so. One article has the engaging title 'I know what I want to teach them and I can't waste time doing what they want to do: Myth and reality of scientific investigation in the junior school classroom' (Young, 1994). A study with elementary pre-service science teachers in Korea revealed a range of concerns that the pre-service teachers had including: learners already knowing 'the answer' to the inquiry; inquiry sometimes producing 'the wrong results' which can be confusing to learners; teachers losing face when the practical work does not work; students making a mess and risking hurting themselves; and the demands on elementary teachers given that they have to teach many subjects (Kim & Tan, 2011).

At its worst, IBSE can make unrealistic demands on pupils and take up inordinate amounts of time. However, it can be motivating for pupils, help them learn science (National Academies of Sciences, Engineering, and Medicine, 2022) and give them a better understanding of how science is undertaken by professional scientists. A systematic review of the outcomes of Independent Research Projects in school science (IRPs) for upper secondary students found that "Benefits were identified in relation to the learning of science ideas, affective responses to science, views of pursuing careers involving science, and development of a range of skills" (Bennett *et al.*, 2018, p. 1755). While the review was restricted to the upper secondary age range, it is worth citing here as it explicitly considered the impact of IRPs for disadvantaged students, concluding:

> One finding of particular interest to emerge from some studies in the USA and the UK was potential benefits to traditionally under-represented groups in science in relation to ethnicity and socio-economic status. In the USA, four studies report improved engagement for such students (Duran et al., 2014; Rivera Maulucci *et al.*, 2014; Sonnert *et al.*, 2013; Yasar & Baker, 2003), with Sikes and Schwartz-Bloom (2009) noting interest declining slightly. In the UK, the British Science Association (2014) found that uptake of IRPs was higher than average for students from lower socio-economic groups. The Nuffield Foundation (2013) reports particular benefits in engagement for students from disadvantaged backgrounds.

(Bennett et al., 2018, p. 1768)

Constructivist approaches

At its simplest, constructivism is a framework for learning within which individuals are envisaged as making (constructing) new knowledge and integrating it with what they already know. In social constructivism, which largely derives from the work of Lev Vygotsky (1896-1934), the emphasis is on the role that other individuals play in enabling the learner to develop knowledge. It was Vygotsky who introduced the notion of the 'zone of proximal development', the idea that, supported appropriately by another, an individual can make progress in learning more than they can on their own – by, metaphorically, as it were, having their hand held by the more knowledgeable learner as they enter this more demanding zone. Jerome Bruner (1915-2016) helped build on Vygotsky's ideas by developing the notion of instructional scaffolding, by which a learner is helped, through a temporary scaffold erected by another, to develop its learning. As the individual's knowledge grows, the scaffold can be dismantled.

Osborne points out that constructivism has its roots in science education in a reaction against two alternative views on learning about science: one, an epistemology based on naïve empiricism (in which it is presumed that we can derive knowledge directly by our observations of the world); the other, "a developmental stage model of cognitive growth which was interpreted as implying deterministic limitations to children's capabilities" (Osborne, 1996, p. 53).

So far there does not seem to be anything controversial but, as with 'inquiry', the issue of constructivism has given rise to battles in science education. The controversy can be illustrated by introducing the notion of 'radical constructivism', a term initiated by the philosopher and psychologist Ernst von Glasersfeld (1927-2010). Von Glasersfeld began from the non-contentious argument (non-contentious among philosophers, that is) that we cannot prove anything for certain about the world outside of our mind because we cannot access that world beyond our experience of it. This is a position worked out by none other than Immanuel Kant (1724-1804) in The Critique of

Pure Reason, which he wrote to address problems that date back, in the Western tradition, to Descartes (1596-1650), who famously started from the point that the one thing of which he could be sure about the world was that he was thinking), and Hume (1711-1776), whose sceptical questioning allowed almost anything to be doubted.

An enthusiastic espousal of von Glaserfeld's point that we each of us cannot prove anything for certain about the world outside of our mind (a point nicely illustrated by many science fiction films of which The Matrix franchise is perhaps the most widely known – would you take the red pill?) is clearly inimical to science and science education. We might as well imagine what the world is like as take the time to investigate it carefully.

Unsurprisingly, scientists take no notice of radical constructivism and few science educators do either, except in the more theoretical of their writings. Nevertheless, critiques by science educators of constructivism (e.g., Osborne, 1996; Matthews, 2002, 2021), while generally careful to distinguish the various categories of constructivism, some of which they identify as pedagogically helpful, can be read as though there was a huge rift on this issue amongst those responsible for curricula, pedagogy and assessment in science education, which there is not.

Constructivism has been important in school science curricula (Guo, 2013) and textbooks (e.g., Andersen, 2018) and is still widely advocated for primary teachers (e.g., Stamp & Preston, 2021). In our experience and more widely (e.g., Beck & Kosnik, 2012; Basturk, 2016), social constructivism (even though not always labelled as such) plays a key role in initial teacher education, as pre-service teachers are encouraged to make scientific predictions, talk about their ideas and then have these scaffolded by their science lecturers/tutors. One study found that the more sophisticated epistemological beliefs pre-service science teachers have, the more constructivist learning environments they prefer (Saylan *et al*., 2016).

A recent example of work in primary science education undertaken within a constructivist framework is provided by Voon *et al.* (2020). An experienced teacher worked with two Year 4 (10 year-olds) classes in Singapore, one with 52 pupils, the other with 57. In lessons on heat transfer, the intended objects of learning were: (a) heat gain/loss results in temperature changes; (b) heat flows from a place with a higher temperature to a place with a lower temperature until both reach the same temperature. Pupils were asked to identify and record the natural phenomena of 'heat transfer' in their daily lives; they then shared these ideas and consolidated them in class. In a lesson, pupils then touched and discussed the temperature of different parts of a pair of hand scissors (with metal blades and a plastic handle) in groups. At the end of the lesson, the pupils were required for an assignment to interview two people outside-of-school (e.g., family members) to gather their opinion on the temperatures of different materials that made up the blade and handle of the scissors. Back in class, the pupils then measured the temperature of the different parts of the scissors using an infrared thermometer, and then had an in-class discussion, facilitated by the teacher. Pupils had completed a worksheet on temperature and heat transfer after they had conducted their interviews; they then corrected their own worksheets after the in-class measurements and discussion.

There are other approaches to learning that relate to constructivism. For example, schema theory is quite widely cited in science education as it is in a number of other subject areas, notably reading. Schemas can be defined as "data structures for representing the generic concepts stored in memory. They exist for generalized concepts underlying objects, situations, events, sequences of events, actions, and sequences of actions" (Rumelhart & Ortony, 1977, p. 101). So, for example, a schema for 'an animal' formed at pre-school level, when 'animal' may be near synonymous with 'non-human mammal', can make it difficult for a learner subsequently to accept that humans, birds, insects and other non-mammalian animals are animals too.

Direct instruction

Often positioned as being the opposite of constructivism or discovery learning (an even more open-ended version of inquiry-based learning where, its opponents sometimes claim, not entirely unfairly, a child on its own is meant to arrive at Newton's Laws of Motion, or whatever, from their classroom activities) is what is sometimes referred to as 'traditional teaching' but is perhaps more usefully characterised as 'direct instruction'. A robust, and highly cited, defence of direct instruction is provided by Kirschner *et al.* (2006), the title of whose article does not mince words: 'Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching'.

There is some imprecision about the language that is used when talking about 'direct instruction', 'explicit instruction' or 'traditional teaching' and its alternatives. As Klahr (2009) points out, when referring to the apparently endless arguments about which type of teaching is best:

> However, these arguments typically fail to establish a common vocabulary to define the essential aspects of the types of instruction being compared. I believe that in order to advance our ability to create effective instructional procedures, our field needs to become much more precise in the terminology it uses to describe instructional contexts and procedures, before moving on to advocacy about curriculum design. In the area of science education, more than others, it is particularly troubling – and ironic – that these debates often abandon one of the foundations of science: the operational definition. But a scientific field cannot advance without clear, unambiguous, and replicable procedures.
>
> (Klahr, 2009, p. 291)

The key difference between direct instruction and its alternatives is nothing to do with the former supposedly being about 'passive learning' and the latter about 'active learning' (neither term having much if any validity) but with the role of the teacher. In direct instruction, the teacher is seen as the expert. The core function of teaching is then to enable the learner to gain the intended knowledge, understanding and skills from the teacher; the thoughts and actions of the learner and their interactions with their peers and learning materials are means to that end. This is the case whether we are talking about the learning of concepts (e.g., the differences between solids, liquids and gases), subject-specific skills (e.g., how to use a quadrat to map vegetation) or generic skills (e.g., how to make a presentation to one's classmates or write up a report of a project).

There is therefore a close link between the arguments for direct instruction and those about powerful knowledge advocated by Young in his aptly titled book Bringing Knowledge Back In (Young, 2008). As is well known, in this book Young argued that the school curriculum must not be based on everyday practical experience but provide learners with access to the specialist knowledge that they need and which cannot be obtained outside of school, for instance from their homes. Young's position is therefore an egalitarian one that seeks to challenge educational disadvantage and reduce educational inequalities:

> The school, for all its tendencies to reproduce the inequalities of an unequal society, is the only institution we have that can, at least in principle, provide every student with access to knowledge. The only alternative to schools for all is to accept that the majority will never have the educational opportunities that the minority has always treated as their right. We

> must respect and value the experience of pupils, but we can never allow them to depend on their experience alone. To do so would leave them (and us) in the position of our Stone Age ancestors, or worse; we would be no different from animals, who have only their experience.
>
> (Young, 2014, p.13)

It is sometimes presumed that direct instruction equates to teaching about concepts, but this is not necessarily the case. For example, Kruit *et al.* (2018) looked at the effects of explicit instruction (as they called it) on the acquisition of pupils' science inquiry skills in grades 5 and 6. They note that such explicit instruction is uncommon since "Skills for scientific inquiry are usually – if at all – taught by teaching methods primarily based on learning by doing" (p. 421). Kruit *et al.* compared the effects of explicit instruction with implicit instruction (where "all aspects of explicit instruction were absent", p. 421) in the primary science topic of heat and temperature. Explicit instruction entailed the teacher clarifying the rationale of inquiry skills, followed by examples and classroom discussions about how to apply the skills. The newly learned skills were then practised with the support of written prompts. For example, in learning how to formulate a research question, the teacher gave explicit instruction about the criteria for formulating research questions with the help of a flow chart. Subsequently, when pupils were asked, for instance, to distinguish between properly and poorly formulated research questions, pupils were reminded of the flow chart. Next, pupils performed a simple whole-task: an authentic, structured inquiry in which they had to apply all skills in an integrated manner. In this task, prompts were explicitly incorporated as well. During the course of the intervention, whole-task inquiries gradually increased in complexity, and the prompts were withdrawn. Finally, pupils performed a scientific inquiry independently. In various post-intervention tests, pupils who had received explicit instruction performed better. Pupil enjoyment did not differ between the two types of instruction.

Talking about science

We have already alluded to the importance of talk for the learning of science when discussing the importance of 'minds-on' thinking for getting the most from practical work and when examining social constructivism. There has long been a literature about the importance of talk for science learning, in primary schools and elsewhere (e.g., Sutton, 1992; Tunnicliffe, 2013). Having said that, it may be worth mentioning the motto of the Royal Society – Nullius in verba (No-one's words) – chosen soon after its founding in 1660. The Royal Society itself explains that this:

> is taken to mean 'take nobody's word for it'. It is an expression of the determination of Fellows to withstand the domination of authority and to verify all statements by an appeal to facts determined by experiment.
>
> (Royal Society, 2020)

Indeed, learning in science in a primary classroom is all about the interplay between four things:

- the material world;
- the understanding of the material world that scientists have;
- the understanding of the material world that the teacher has;
- the understanding of the material world that the pupils have.

Direct access to the material world for the pupils in a primary classroom is fairly modest, being that provided by their powers of observation and the equipment and materials that they use in practical work. Language plays an important role in helping pupils learn directly from the material world;

indeed, it constitutes almost the only way that they learn from their teacher, their peers and (principally via their teacher and any science books they consult) from scientists.

In this section we concentrate on the value of talking for learning. Of course, pupils need not only to talk about science – they need to read about it, to listen about it and to write about it. These are the four domains that constitute literacy; in addition, nowadays we are more aware than ever that language does not only operate through words. We live in a world where multimodal learning is increasingly important (Jewitt, 2005; Crescenzi-Lanna, 2020); indeed, scientists themselves make tremendous use of photographs, diagrams, graphs and other modes of language (e.g., Roth & Bowen, 2001). Mayer's cognitive theory of multimedia learning is based on three principles: a dual-channel assumption (namely, that the human information processing system has one channel for auditory/verbal processing and one for pictorial processing; each channel has a limited capacity for processing; and learning requires active processing (Mayer, 2005).

Some of the literature about language in science is about the problems with scientific language that learners can experience; some of it is about how learners can use language to help them learn science. Sutton makes a useful distinction "between the learner's experience of language as an interpretive system, actively used for generating new understanding, and of language as a labelling system for transmitting established information" (Sutton, 1996, p. 1). Sutton points out that examples of language for the transmission of established information include such statements as 'Copper turns black when heated' and 'Air molecules are in constant motion', whereas the musings we find in William Harvey's 1628 writing about blood ('I began to think whether there might not be a motion, as it were, in a circle') and Watson's and Crick's writing about DNA ('It has not escaped our notice that the pairing we have postulated immediately suggests a possible copying mechanism for the genetic material') are classic instances of language operating to help make sense of new experiences and ideas.

Much of school science is about helping pupils use language interpretatively so that they build up their knowledge of established science. We want primary children to learn what scientists mean by food – so that it is not really appropriate to talk about plants taking in food. However, pupils need to be encouraged to talk about plant nutrition using whatever language they have before they can come to a scientific understanding of the concepts. Similarly, it is fine for pupils to come to an understanding of camouflage by talking about animals 'wanting' to be invisible – but eventually we want them to realise that camouflage in animals does not require this sort of conscious thought on the animals' part.

In a detailed ethnographic study of two effective Australian primary science teachers, Fitzgerald (2012) collected data on Deanne's class of Year 7 (11-12 year-olds) pupils (all names are pseudonyms). In one unit consisting of nine lessons, Deanne wanted the children to learn about scientific terminology in chemistry. Here is an extract from a conversation among one group of children in the first lesson of the unit:

Mark	This is absolutely custard powder.
Anna	How do you know?
Natalie	This is custard powder. Substance turns gooey, yellow with orange spots.
Evan	It's glunky

Anna It's not custard powder because it's orange.

Mark It is.

Anna Custard powder is pale yellow.

Natalie It doesn't matter.

Anna It does matter.

Natalie Can I just say, it doesn't matter what it is. It matters what it looks like.

Evan What are we going to write?

Natalie The substance turns sticky, yellow and thick with orange spots.

Mark I'll give you any money if that's not custard powder.

[Teacher joins group]

Natalie That looks a lot like custard powder.

TeacherWhy do you say that?

Mark Because custard powder goes the same yellow.

Anna But custard powder is a very pale yellow.

TeacherBut that's not pale yellow.

Anna: I know. That's why I don't think it's custard powder.

(Fitzgerald, 2012, p. 99)

As Deanne introduced chemistry terminology throughout the unit, the pupils changed the way that they talked. Here is an extract from a conversation among the same group of children in the eighth lesson of the unit:

Anna Wow! [The custard powder] is definitely dissolving.

Yvette And that would be soluble.

Anna So do we all agree it's soluble? It does dissolve because it was in a big clump before.

Mark I think the custard powder is soluble.

Anna Why? Wy?

Mark What do you mean why?

Anna Why is it soluble?

Mark Because it looks soluble.

Anna Because it looks soluble? It's soluble because it dissolves.

Mark I suggest we leave it so the custard powder can settle.

Yvette	Do you think its [sic] see through? Oh it is. Because you can see the spoon through it.

Mark	No, leave it Yvette. Then you can properly see if it's a suspension. If you wait, you'll see what I mean. See there is already a gap [indicating to the cup and the slowly settling custard particles].

Yvette	Do you think it's a solution?

Anna	It's not a solution.

Yvette	Yes, it is.

Anna	No, it's not clear.

Mark	It is clear. Look. It's translucent.

Yvette	No, it's a solution because the solute dissolves in the solvent.

Anna	If we let it settle. So right now, it's only translucent when it settles?

Yvette	No, it's still translucent.

Mark	Don't pick it up.

Yvette	But that one [custard powder and water] does dissolve and that's why it's a solution.

Mark	I think it's actually, it's soluble, but …

Anna	Look! There's powder, you can see it.

Evan	Exactly, so it's a suspension.

(Fitzgerald, 2012, pp. 100-101)

Tackling misconceptions

The term 'misconception' is not liked by all science educators, with some arguing that it patronises children and that we should therefore talk of 'alternative conceptions' or simply 'children's ideas'. Whatever language is used, though, the reality is that children, including primary-aged children, have many ideas about the material world that conflict with standard scientific accounts (e.g., Allen, 2020). In that sense, it makes sense to talk about children's misconceptions. As Harlen and Qualter put it:

> The studies of Piaget in the first part of the twentieth century revealed that not only were young children eager to interact with the things in their environment, but as a result they developed ideas about the world around. Researchers who replicated these studies in various parts of the world found remarkably similar ideas arising in quite different contexts. It was soon realised that the existence of these ideas, which were often in conflict with the scientific understanding of events and relationships, had an impact on children's learning in science.

(Harlen & Qualter, 2018, pp. 102-103)

In the UK, the SPACE (Science Processes And Concepts Exploration) project, undertaken from 1987 to 1990, studied primary children's ideas across the whole curriculum, resulting in ten published research reports, a number of articles (e.g., Harlen, 1992) and other outputs (e.g., Harlen & Qualter, 2018). This and other related projects showed the value of attending carefully to children's ideas:

> The overwhelming conclusion from studying children's ideas on a wide range of scientific topics is that they are the product not of childish imagination, but of reasoning. The gap between the children's ideas and accepted scientific ones may arise because of the children's limited experience (e.g., they assume that all wood floats because they have never seen any that does not) or from immature reasoning and selective use of evidence.
>
> (Harlen, 1992, p. 499)

To give just one example of a widespread misconception that the SPACE research revealed:

> It is very common for children (and indeed some adults) to regard the eye as an active agent in seeing, rather than being a receiver of light from a source or reflecting surface. This view corresponds with the feeling of moving the eye to 'look' from one place to another and with the words we use such as 'casting a glance' or 'looking hard'.
>
> (Harlen & Qualter, 2018, p. 120)

Ofsted points out that when dealing with misconceptions in science, lessons from the history of science can be helpful as these can help "pupils to see how their initial conceptions mirror those of early scientists" (Ofsted, 2021).

The fundamental thinking behind taking account of pupil misconceptions in science is that if one does not, pupils are likely to stick with their existing ways of understanding the world (e.g., Duit & Treagust, 2003), even if scientists hold that these are mistaken. To an evolutionary biologist, this makes a lot of sense. Our everyday understandings of the natural world have served us pretty well for millions of years. The science that gets taught in schools is a comparatively recent phenomenon, rarely dating back more than a few hundred years. Take Newton's First Law, often taught in primary science in the context of forces, namely that an object continues in a state of uniform motion unless acted upon by a net force. It is difficult to overstate how useful Newton's First (and other two) Laws are to physicists and engineers. However, to the proverbial person on a Clapham omnibus, Aristotelian notions of movement – that the natural state of a moving object is to slow down – work very effectively. Indeed, it is difficult in a primary school to produce much convincing empirical evidence in favour of Newton's First Law (a secondary physics lab may have a linear air track).

Pupil misconceptions can be held very strongly. Allen and Coole (2012) cite the example of Jennifer who, when asked to take the temperature readings of two thermometers, one in a lagged beaker of cold water and one in an unlagged beaker of cold water, initially reported that the temperature of the water in the lagged beaker was 5 °C higher than that in the other beaker. After her teacher asked Jennifer to take the readings again very carefully, she correctly stated that the temperatures were equal. As Allen and Coole go on to say:

> It is possible that Jennifer had a strong expectation that the lagged beaker would actively 'warm' the water, and so in her mind all that was required was a quick glance at both thermometers to confirm that was indeed the case. She had either made an innocent, careless mistake (which could have been a random error, or unconsciously driven by her preconception) or actually read the thermometers correctly but then purposely chose to ignore the readings and gave the teacher what she thought to be the 'right answer'.

(Allen & Coole, 2012, pp. 388-389)

Harlen and Qualter (2018) suggest various classroom approaches that a science teacher might take to address pupil misconceptions and enhance conceptual understanding, including the following:

- Provide experiences to show that some things can happen that contradict pupil conceptions – e.g., that heavy objects can float in water.
- Scaffold the introduction of alternative ideas that align with scientific thinking.
- Encourage close observation – e.g., does the shape of an object as well as its mass affect whether it can float in water?
- Help pupils develop inquiry skills.
- Refer to other contexts in which the same idea is applicable – e.g., is there something vibrating in a wind instrument that produces sound, as the vibration of a drum skin does?

As will be apparent, there are links between tackling misconceptions, encouraging talk, using inquiry and adopting a constructivist approach. In a study to address Year 6 pupil misconceptions about the phases of the Moon, Mohd Radzi *et al.* (2017) compared the effectiveness of an inquiry-discovery teaching approach and a more traditional approach. Pupils were randomly assigned to one of the two groups; the pupils in the inquiry-discovery group were found to have superior knowledge and to hold fewer misconceptions.

Cross-curricular learning

There is a widespread presumption that pupils would benefit from science being taught using cross-curricular approaches to a greater extent than it usually is. As Bianchi and Thompson put it in an *Association for Science Education Guide to Primary Science Education*:

> Taking a cross-curricular approach to teaching and learning provides opportunities that have personal relevance for children and connects to their questions and their everyday lives. Approaches from infusing literacy, numeracy and ICT skills, through to using a personalised topic approach, give children the chance to see the links between science and different areas of their learning and enhances their engagement and motivation for science.

(Bianchi & Thompson, 2011, p. 53)

In most primary schools, pupils are taught science by the same teacher who teaches them other subjects. This clearly facilitates cross-curricular learning compared to a situation where different subjects are taught by different specialist teachers. Links between mathematics and science are clear, as are links between science and languages but recent years has shown a rapid growth in interest about the relationship between the arts and the sciences and there are also clear links between science and other subjects, including geography (rocks, soil, the environment in general), history (the history of science, the importance of new technologies for both political and social history) and Religious Education (causes célèbres like evolution but also more general issues to do with revelation and the nature of authority and whether there is purpose in the world).

The literature generally concludes that less is made of cross-curricular learning in primary science than might be the case. There are a number of reasons for this, including the fact that the curriculum is often defined by subjects (as is the case in England and Wales since the advent of the National Curriculum) and this may be reinforced by formal assessment arrangements. In a study of elementary science teachers in the US, the authors concluded that the teachers "lacked a conceptual connection to integration, showed contradictions in the importance placed on hands-on experiences, used measurement as the primary interdisciplinary connection between mathematics and science, and did not use instructional strategies designed specifically for nonfiction/expository text" (Douville *et al.*, 2003, p. 388). This study ended with a plea, as do many other studies on cross-curricular teaching, for more teacher professional development.

Often, there is enthusiasm for encouraging children to write more creatively in primary science than they generally do. For example, Weiss-Magasic, a US elementary teacher, writes:

> In elementary school, there's always a reason to celebrate, but in the upper grades, class time is more serious. I occasionally abandon that seriousness to celebrate a calendar holiday or one that I make up. For example, I stage a celebration for Charles Darwin during February's winter doldrums and have students honor him through letters, poems, songs, and so on.
>
> On Darwin's birthday, February 12, students circle their desks, creating a "coffeehouse" (without food or drink – these are prohibited in lab classrooms). Each student presents an ode to Darwin that must incorporate something personal and scientific about him. I instruct students to avoid obvious facts and present something surprising and insightful that shows comprehension of Darwin not just as an icon but as a person. I allow shy students to turn theirs in before class and alternate my dramatic renditions of their work with other students' performances.

(Weiss-Magasic, 2012, p. 42)

In another study that drew on a cross-curricular approach to primary science, an example is provided from Romania of how this enriched children's learning:

The teacher told a story to her class of children aged 5 and 6 about an ant who fell into the river. A dove flying by wanted to help the ant. By providing an inquiry-based problem that had more than one solution and by giving children autonomy to come up with their own ideas, the children were able to plan their investigations and showed creativity in generating their own ideas about which materials to use and how to test them, using their imagination and making connections with prior experiences. The children discussed natural materials in the forest the dove might use to help keep the ant afloat. A variety of materials was made available, including nuts, feathers, wooden sticks, leaves, little stones, acorns, pieces of bark, fir cones. Each group discussed their own predictions about the materials they thought most suitable to save the ant. They were given small containers with water to test their ideas about which materials in the forest could be used as little 'boats' for the ant. Children were able to record and communicate findings in their own ways. Children shared and evaluated their findings, drawing on evidence from their observations, to justify conclusions about whether this object would be appropriate to help save the ant.

(Stylianidou, 2014, p. 11)

Acar *et al.* (2018) undertook a study on the consequences of integrated STEM (Science, Technology, Engineering and Mathematics) professional development for elementary teachers in Turkey on their Grade 4 pupils' learning in mathematics and science. Compared to control classes, the experimental classes achieved higher on achievement tests for both mathematics and science, though the experimental classes were taught by the researcher, whereas the control classes were taught by the regular teachers and, as is not uncommon in evaluation of cross-curricular teaching, there were differences in the teaching that were not only to do with cross-curricular matters, such as greater use of project work, group work and presentations in the experimental classes. Nevertheless, some of the qualitative data suggest that the cross-curricular work was appreciated by the pupils with, for example, one pupil stating:

> I'd want future courses to be similar. We learned both science and mathematics. Meanwhile Turkish also came into play, such as reading. Sometimes social studies also came into play. Therefore we can do it, and it's fun. (Kemal Mert)
>
> (Acar *et al.*, 2018, p. 510)

A number of countries have moved to explore the possibilities of a 'STEAM' curriculum instead of a STEM one (e.g., Kwan & Wong, 2021), adding 'Arts' to 'STEM', and there has been considerable enthusiasm about this among science educators (Colucci-Gray *et al.*, 2017; Braund & Reiss, 2019). Braund and Reiss propose four main premises in their argument that science is made more complete by its relationship with the arts. They use these to lead to a fifth premise, establishing a case for the arts to enhance the ways in which science courses and teaching methods might change to make science learning more authentic and engaging:

1. The subject boundaries premise: Divisions between curriculum areas (school subjects) run counter to the life experiences of learners of all ages.
2. The cognitive premise: The work of science needs creative as well as critical thinking to allow discourses that empower and fuel discovery and innovation and allow risk-taking.
3. The neuroscience premise: Thinking in science is stimulated by artistic activity.
4. The collaborative, economic premise: Collaboration between arts and sciences and vice versa is at the heart of the modern economy.
5. The pedagogical premise: The final justification is embedded in science education: organising curricula to accommodate science and arts and drawing on pedagogy normally associated with the arts offer fruitful ways to engage learners in school science and help them learn and to help prevent young people turning away from science.

(Braund & Reiss, 2019, p. 222)

Play

Much of the literature on play and science education focuses on the pre-primary phase, though authors often argue for the value of play for all phases of learning. Indeed, scientists not infrequently talk about the benefits of their 'playing' with equipment or data as a way of familiarising themselves with these (cf. Perkel, 2019). Mention is sometimes made of the way that Watson and Crick played around with their DNA models. One of us was taught how to teach physics during their initial teacher education by a charismatic physics educator who collected physics toys all her professional life, and used them in her teaching. In general, play entails a range of activities that are intrinsically rewarding and are performed, often with others, for pleasure. The extent to which such activities are structured varies greatly, with unstructured play being more common among younger children and in informal settings. As a species, humans are characterised by extreme neoteny, so we play for years longer than other mammals.

There is a large literature on the benefits of play for learning, even though play and learning are often expressed as alternatives (Pyle & Danniels, 2017). A distinction is sometimes made between 'free play' (play that is child-directed and voluntary) and 'adult-guided play', though as a systematic review noted "Numerous taxonomies of play are in circulation" (Bubikova-Moan *et al.*, 2019, p. 777).

In general, play has been found to have academic benefits and to enhance the development of social-emotional skills, including self-regulation skills. Marilyn Fleer (e.g., Fleer *et al.*, 2022) has long argued that play-based settings can enable greater learning of primary science concepts.

But play is becoming less frequent in schools. Bailey *et al.* (2019) bemoan the fact that pressures to improve attainment can lead to hyper-vigilance about children's behaviour in the classroom, inhibiting children's abilities to build and practise self-regulation skills. In a school setting, play is often regarded as frivolous and a waste of time unless it occurs in 'break time'. 'Fiddling with equipment' in science classes is often equated by teachers as 'messing about' in the pejorative sense (e.g., Doveston & Keenaghan, 2006), rather than as a necessary phase of familiarisation.

The value for pupils' science of their playing with equipment has been recognised (e.g., Tifi *et al.*, 2006) and it is often argued that play is of particular value for young children as they start to learn science. A typical example for the journal Science and Children, under the heading 'Scientists at play', is as follows:

> Children are scientists at play. Watch as they bake mudpies or construct worm playgrounds. Listen to them. Amidst giggling you may hear an exchange of observations or well-thought-out theories. Playfully explore with them and you may be astonished at how their attention far outlasts yours. Children use many of the same science process skills as adult scientists. Similarly, they develop ideas about our world based on experiences with real things.
>
> (Ross, 1997, p. 35)

In an article titled 'Play as the learning medium for future scientists, mathematicians, and engineers', Bergen advances a number of mechanisms that might account for the educational benefits of play:

1. Play serves as a channel of communication for children who are not always articulate in other ways.
2. Play enables them to examine materials and try techniques in artistic and creative endeavors.
3. Play helps them convey ideas and accomplish goals before their language skills are fully developed.
4. Play "substance" provides a filter that allows them to take risks without concern for world realities.
5. Play allows them to feel powerful in transmitting forceful ideas and producing exciting effects.
6. Play promotes an optimum learning environment within which they can function and flourish naturally.

(Bergen, 2009, p. 416)

One formalised form of play that occurs in science lessons is when teachers use role play in their teaching, part of a growing literature on the potential for drama to be used effectively in science classrooms (e.g., Ødegaard, 2003; Darlington, 2010; Braund, 2015). Aubusson and Fogwill provide a theoretical underpinning for the value of role play in science education. They see role play as helping learners to create their own mental models; specifically, they argue that role play can serve as analogical modelling. They note that all models are flawed and so break down. Accordingly: "We assert that an analogy is successful not when it most accurately portrays ideas per se but when it promotes conversation, central to producing, evaluating and modifying the analogy, that helps students to clarify and to improve their scientific understanding" (Aubusson & Fogwill, 2006, p. 94).

Cakici and Bayir (2012) report on the benefits of using role play on the understanding of 10-11 year-old children about NOS. The children were taught a unit about the work of Isaac Newton and Marie Curie and helped to reflect on NOS through being asked such questions as 'What do you think of how Newton showed the existence of gravity?', 'Did he show it in his hands, in a photograph or on the table?', 'Where is the gravity in this class?' and 'How objective is Newton?'. They also undertook and performed a number of role-play activities. After the unit, the children showed substantially better understanding of such ideas as: science knowledge can change over time; a key feature of science is the gathering of empirical data; scientists use multiple methods to gather data; and the environment and society that scientists live in may affect their views or scientific studies.

Lessons from neuroscience

As its name suggests, neuroscience is the scientific study of the nervous system. The last couple of decades have seen an explosion of claims as to what neuroscience has to say to education. Perhaps unsurprisingly, as is often the case in any new intellectual movement that attempts to colonise a well-established field (think of the 'contributions' of IQ testing and genetics for education), there is initially much hype and many initially enthusiastic professionals subsequently retreat, having had their fingers burnt. In the case of educational neuroscience, examples of what turned out to be fads include crude understandings of learning styles (Coffield *et al.*, 2014) and the difference between left-brain and right-brain thinking (Lindell & Kidd, 2011). Indeed, a recent review on cognitive science in the classroom concluded that "There are large disconnects between the evidence-base for basic cognitive science and applied cognitive science. Applied cognitive science is far more limited and provides a less positive, and more complex, picture than the basic science" (Perry *et al.*, 2021, p. 260).

Nevertheless, education is fundamentally about learning, so it is hardly surprising that the discipline most concerned with how the brain works should have something to say to education. There are reputable centres and journals that specialise in educational neuroscience and a growing peer-reviewed literature. Indeed, by taking seriously the biology of the human nervous system, neuroscience can be seen as trying to move beyond earlier attempts to explain learning – such as the information processing model of learning where the brain was in a sense considered as a black box, and analogised to be like a computer (hence the focus on memory, encoding and storage).

Educational neuroscience has its roots in psychology and one example of a psychological principle that can now be considered part of neuroscience and has implications for learning is the concept of 'working memory', a term that dates back to the 1960s. Working memory is also known as short-term memory (though some distinguish between the two) which, as the term suggests, is a transient memory, one that sits between sensory memory (which has a very large capacity but can only hold memories for milliseconds) and long-term memory (which is lifelong, though long-term memories fade if not rehearsed and can be lost). As a rule of thumb, working memory holds memories for several seconds and cannot hold more than about five to seven 'items' of information – whether the items are numbers, letters, words, images of objects, or whatever.

The immediate implication of working memory for a teacher is to avoid filling learners' working memories up with irrelevant information or with too much relevant information (cf. Reid, 2009). Another implication is that teachers should employ strategies to maximise the likelihood that desired memories (what we want pupils to learn) are moved into the long-term memory, where they are subsequently rehearsed. None of this is likely to come as a great surprise to an experienced teacher, whether of primary science or any other subject at any phase of education, but it may be useful to have the underpinning mechanism made explicit. Furthermore, there have been claims that 'brain-based teaching' can lead to increases in working memory and result in enhanced academic motivation (Adel & Mourad, 2019).

Allied to working memory theory is variation theory. Variation theory attempts to help explain why different learners learn so very differently from the same teaching (Bussey *et al.*, 2013). Obviously, there is a range of reasons for this including the extent of prior knowledge, and learner differences in motivation. Variation theory points out that any phenomenon presents a large number of features to a learner but (and here is the similarity to working memory theory) a learner can only pay attention to a relatively small number of these features. It is therefore unsurprising that even if two learners come to a phenomenon with similar motivation and background knowledge, they may experience and learn very differently. For the teacher, the key point to bear in mind is that while they (the teacher) may find it trivial to understand what is going on when, for instance, students are asked to compare the height to which balls bounce on different surfaces, the quick demonstration they provide may lead to learners noticing very different things and failing to notice certain crucial things (e.g., that the ball must be dropped – and not tossed – from the same height irrespective of the surface).

More generally, Goswami points out that "Learning by the brain depends on the development of multi-sensory networks of neurons distributed across the entire brain. For example, a concept in science may depend on neurons being simultaneously active in visual, spatial, memory, deductive and kinaesthetic regions, in both brain hemispheres" (2015, p. 25). Harlen concluded, from her review of the literature on the implications of neuroscience for primary science:

> Perhaps not surprisingly, this evidence found at the microscopic level of brain cells confirms what is found at the macroscopic level of studying how pupils respond to educational stimuli. Both support the assertion that pupils should be provided with experiences that:
>
> - interest and engage children – are seen by them as relevant and appealing
> - build upon their previous experience, allowing some repetition to consolidate and apply learning
> - provide challenge within the reach of children so that they experience pleasure in learning
> - engage the emotions by making learning science exciting.
>
> (Harlen, 2011, p. 40)

Harlen also made some more tentative conclusions that again will not surprise experienced science educators and reinforce a number of conclusions drawn in earlier sections:

> There is also support for teaching methods that:

- encourage talk, argumentation and exchange of ideas among pupils
- enable active investigation and the use of the senses
- provide practice in using skills and applying ideas
- create habits of using representations and keeping notes to aid memory
- ensure that pupils understanding [sic] their goals and how to assess achievement (formative assessment)

- encourage reflection on what they are learning and how they are learning, and awareness of what they don't understand
- promote the use of scientific ideas, once formed, in preference to intuitive naïve theories.

(Harlen, 2011, pp. 40-41)

A related point is that it is increasingly acknowledged that knowledge is 'situated'. Instead of viewing the learner simply as accessing previously stored knowledge, whenever the situation so requires, a learner's manifestation of their knowledge is seen as being intimately connected to the context (hence the theory of situated cognition). The child who seems to know little of food chains and food webs in the classroom apparently metamorphoses into one who is able on a field trip to talk in detail about what three-spined sticklebacks catch and eat and how they behave in various circumstances. For the teacher this does not mean, of course, that knowledge can never be transferred from one context to another but that learners benefit from being helped to make such transfers: "You remember how last lesson we talked about all matter being made up of particles and used that scientific model to help explain what is going on when water evaporates? Today we are going to see if the same model can be used to explain what happens when liquid water freezes to become ice, and when water vapour condenses to become liquid water".

Learning out-of-doors

Finally, there has recently been a growing interest in the affordances and value of learning out-of-doors for primary science, in part fuelled by an increasing acknowledgement of what the Anthropocene / Capitalocene is doing for both the natural environment and people's health and wellbeing (cf. Louv, 2005). The mechanisms that are likely to underlie any benefits of out-of-doors learning are partly to do with learner motivation and partly to do with place-based education, where there is a contextual specificity and tangibility to learning.

Out-of-doors learning need not necessarily take place away from school premises. A study of the 'Gardens for Bellies' Danish school garden programme revealed benefits for both primary pupils and secondary students with regards to wellbeing and self-esteem (Dyg & Wistoft, 2018). In a review of the progress of the Royal Horticultural Society's (RHS) Campaign for School Gardening, Passy (2014) undertook fieldwork in ten primary schools (seven urban and three rural) that had participated in the campaign. Using a framework of possible cognitive, affective, behavioural and physical, and interpersonal and social gains, a number of benefits were identified, even though there was a divide between teachers who were willing to use the garden in their teaching and those who were not. Examples of benefits were:

> You work better as a team in the garden; in the classroom you can't work together as much because you're sitting in rows and you can't be as friendly with people. Often in the garden it changes you. You work with more people in the garden, not just your friends. It will make you more friendly with them, it does kind of help you make more friends. (pupil)

(Passy, 2014, p. 32)

> ... having this wonderful space is something that enables us to be as creative as we possibly can be in order to, to get those children [who lack confidence] to a point where they have good self-esteem, they have skills, they want to learn more. Because most of the time they do, once they've started on something, they want to know more. And their questioning skills … they romp ahead with questioning skills because they want to find out more. Even if they're just asking themselves, I wonder what's going to happen to that, how long it will take

> to grow, you know, when will we get the flowers, when can we harvest this. (deputy head teacher)
>
> (Passy, 2014, p. 34)

Moving beyond the confines of school gardens, Waite (2011) has a raft of suggestions for nature activities outside the classroom and Coll and Coll (2019) focus specifically on how science learning can be enhanced through learning experiences outside of school. Recent years have also seen growing enthusiasm for Forest Schools (e.g., Knight 2009). Knight points out that while Forest Schools ideally take place in woods (there is a clue in their name …), they can take place elsewhere. In one of her case study schools, Forest School actually took place in the school grounds – the important thing was that Forest School rules apply there. As she puts it:

> The learning is play-based and, as far as possible, child-initiated and child-led. There are no time constraints, and risk-taking is facilitated. Forest School is about an internal process of holistic development, something that is difficult to achieve in a busy classroom, indoor or out.
>
> (Knight, 2009, p. 17)

To date, evaluations of the learning benefits of out-of-doors activities have perhaps lagged behind enthusiasm for them. Sheldrake *et al.* (2019) looked at the impact of Wildlife Trust events in England that involved primary-aged children learning about nature while out of doors. The research surveyed children before and after they undertook Wildlife Trust events, and also undertook observations of and interviews with children, teachers and Wildlife Trust practitioners who delivered the events. Increases were revealed over time for the children's subjective well-being, nature connection and pro-environmental values.

These findings were supported through the children's reflections on their own experiences, and through observations. Children's enjoyment levels were seen to be high; their motivation and engagement were high; and they exhibited curiosity, active observation and engagement with nature.

References

Abrahams, I. (2009). Does practical work really motivate? A study of the affective value of practical work in secondary school science. *International Journal of Science Education*, *31*(17), 2335-2353.

Abrahams, I. & Reiss, M. (2012a). Evolution. In: *The Routledge International Handbook of Learning*, Jarvis, P. with Watts, M. (Eds), pp. 411-418. Abingdon: Routledge.

Abrahams, I. & Reiss, M. J. (2012b). Practical work: Its effectiveness in primary and secondary schools in England. *Journal of Research in Science Teaching*, *49*, 1035-1055.

Abrahams, I., Reiss, M. J. & Sharpe, R. (2014). The impact of the 'Getting Practical: Improving Practical Work in Science' continuing professional development programme on teachers' ideas and practice in science practical work. *Research in Science & Technological Education, 32*, 263-280.

Acar, D., Tertemiz, N. & Taşdemir, A. (2018). The effects of STEM training on the academic achievement of 4th graders in science and mathematics and their views on STEM training. *International Electronic Journal of Elementary Education*, *10*(4), 505-513.

Adel, M. E. & Mourad, A. E. S. (2019). Effect of a brain-based learning program on working memory and academic motivation among tenth grade Omanis students. *International Journal of Psycho-Educational Sciences*, *8*(1), 42-50.
Allen, M. (2020). *Misconceptions in Primary Science, 3rd end*. London: Open University Press.
Allen, M. & Coole, H. (2012). Experimenter confirmation bias and the correction of science misconceptions. *Journal of Science Teacher Education*, *23*, 387-405.
Andersen, K. N. (2018). Evaluation of school tasks in the light of sustainability education: Textbook research in science education in Luxembourgish primary schools. *Environmental Education Research*, *24*(9), 1301-1319.
Aubusson, P. J. & Fogwill, S. (2006). Role play as analogical modelling in science. In *Metaphor and Analogy in Science Education*, Aubusson, P. J., Harrison, A. G. & Ritchie, S. M. (Eds), pp. 93-104. Dordrecht: Springer.
Bailey, R., Meland, E. A., Brion-Meisels, G. & Jones, S. M. (2019). Getting developmental science back into schools: Can what we know about self-regulation help change how we think about "no excuses"? *Frontiers in Psychology*, *10*, 1885.
Basturk, S. (2016). Primary pre-service teachers' perspectives on constructivism and its implementation in the schools. *Universal Journal of Educational Research*, *4*(4), 904-912.
Beck, C. & Kosnik, C. (2012). *Innovations in Teacher Education: A Social Constructivist Approach*. New York: SUNY Press.
Bennett, J. (2003). *Teaching and Learning Science: A Guide to Recent Research and its Applications.* London: Continuum.
Bennett, J., Dunlop, L., Knox, K. J., Reiss, M. J. & Torrance-Jenkins, R. (2018). Practical Independent Research Projects in science: A synthesis and evaluation of the evidence of impact on high school students. *International Journal of Science Education*, *40*, 1755-1773.
Bergen, D. (2009). Play as the learning medium for future scientists, mathematicians, and engineers. *American Journal of Play*, *1*(4), 413-428.
Bianchi, L. & Thompson, P. (2011). Science within cross-curricular approaches. In *ASE Guide to Primary Science Education, new edn*, Harlen, W. (Ed.). Hatfield: Association for Science Education, pp. 53-60.
Bianchi, L., Whittaker, C. & Poole, A. (2021). *The 10 Key Issues with Children's Learning in Primary Science in England*. Manchester: University of Manchester & London: The Ogden Trust. https://www.scienceacrossthecity.co.uk/wp-content/uploads/2021/03/3634_Childrens_Learning_in_Primary_Science_Report_2020_v8.pdf.
Braund, M. (2015). Drama and learning science: an empty space? *British Educational Research Journal*, *41*(1), 102-121.
Braund, M. & Reiss, M. J. (2019) The 'Great Divide': How the arts contribute to science and science education. *Canadian Journal of Science, Mathematics and Technology Education*, *19*(3), 219-236.
British Science Association (2014). *CREST Evaluation Summary.* Unpublished internal document.
Bubikova-Moan, J., Hjetland, H. N. & Wollscheid, S. (2019). ECE teachers' views on play-based learning: A systematic review. *European Early Childhood Education Research Journal*, *27*(6), 776-800.

Bussey, T. J., Orgill, M. & Crippen, K. J. (2013). Variation theory: A theory of learning and a useful theoretical framework for chemical education research. *Chemistry Education Research and Practice, 14*(1), 9-22.

Cakici, Y. & Bayir, E. (2012). Developing children's views of the Nature of Science through role play. *International Journal of Science Education*, *34*(7), 1075-1091.

Christodoulou, D. (2014). *Seven Myths about Education*. London: Routledge.

Coffield, F., Moseley, D., Hall, E., & Ecclestone, K. (2004). *Should We be Using Learning Styles? What Research has to Say to Practice.* London: Learning & Skills Research Centre

Coll, S. D. & Coll, R. K. (2019). *Enhancing Science Learning Through Learning Experiences Outside School (LEOS): How to Learn Better During Visits to Museums, Science Centers, and Science Fieldtrips*. Leiden: Brill Sense.

Colucci-Gray, L., Burnard, P., Cooke, C. F., Davies, R., Gray, D. S. & Trowsdale, J. (2017). *Reviewing the Potential and Challenges of Developing STEAM Education through Creative Pedagogies for 21st Learning: How can School Curricula be Broadened towards a more Responsive, Dynamic, and Inclusive Form of Education?* London: British Educational Research Association.

Correia, M. & Freire, A. M. (2016). The influence of an in-service programme on primary teachers conceptions about practical work. *Revista Electrónica Interuniversitaria de Formación del Profesorado*, *19*(2), 259-271.

Crescenzi-Lanna, L. (2020). Multimodal learning analytics research with young children: A systematic review. *British Journal of Educational Technology*, *51*(5), 1485-1504.

Darlington, H. (2010). Teaching secondary school science through drama. *School Science Review, 91*(337), 109-113.

Department for Education (2015). *National Curriculum in England: Science Programmes of Study*. https://www.gov.uk/government/publications/national-curriculum-in-england-science-programmes-of-study/national-curriculum-in-england-science-programmes-of-study.

di Fuccia, D., Witteck, T., Markic, S. & Eilks, I. (2012). Trends in practical work in German science education. *Eurasia Journal of Mathematics, Science and Technology Education, 8*(1), 59-72.

Douville, P., Pugalee, D. K. & Wallace, J. D. (2003). Examining instructional practices of elementary science teachers for mathematics and literacy integration. *School Science and Mathematics*, *103*(8), 388-396.

Doveston, M. & Keenaghan, M. (2006). Growing talent for inclusion: Using an appreciative inquiry approach into investigating classroom dynamics. *Journal of Research in Special Educational Needs*, *6*(3), 153-165.

Duit, R. & Treagust, D. (2003). Conceptual change: A powerful framework for improving science teaching and learning. *International Journal of Science Education*, *25*, 671-688.

Duran, M., Höft, M., Lawson, D. B., Medjahed, B. & Orady, E. A. (2014). Urban high school students' IT/STEM learning: Findings from a collaborative inquiry- and design-based afterschool program. *Journal of Science Education and Technology*, *23*(1), 116-137.

Dyg, P. M. & Wistoft, K. (2018). Wellbeing in school gardens – the case of the Gardens for Bellies food and environmental education program. *Environmental Education Research*, *24*(8), 1177-1191.

Fitzgerald, A. (2012). *Science in Primary Schools: Examining the Practices of Effective Primary Science Teachers*. Rotterdam: Sense.

Fleer, M., Fragkiadaki, G. & Rai, P. (2022). The place of theoretical thinking in professional development: Bringing science concepts into play practice. *Learning, Culture and Social Interaction*, 32, 100591. https://doi.org/10.1016/j.lcsi.2021.100591.

Gopnik, A. (2009). *The Philosophical Baby: What Children's Minds tell us about Truth, Love & the Meaning of Life.* New York: Random House.

Goswami, U. (2008). *Cognitive Development: The Learning Brain*. Hove: Psychology Press.

Goswami, U. (2015). *Children's Cognitive Development and Learning*. Cambridge: Cambridge Primary Review Trust.

Guo, C. J. (2013). Issues in science learning: An international perspective. In *Handbook of Research on Science Education*, Abell, S. K. & Lederman, N. G. (Eds), pp. 241-270. Mahwah, NJ: Lawrence Erlbaum.

Harlen, W. (1992). Research and the development of science in the primary school. *International Journal of Science Education*, *14*(5), 491-503.

Harlen, W. & Qualter, A. (2018). *The Teaching of Science in Primary Schools, 7th edn*. London: Routledge.

Hodson, D. (1990). A critical look at practical work in school science. *School Science Review*, *70*(256), 33-40.

Jenkins, E. W. (1995). Central policy and teacher response? Scientific investigation in the national curriculum of England and Wales. *International Journal of Science Education*, *17*(4), 471-480,

Jewitt, C. (2005). *Technology, Literacy, Learning: A Multimodal Approach*. London: Routledge.

Johnston, J. (2005). *Early Explorations in Science, 2nd edn*. Maidenhead: Open University Press.

Johnston, J. (2014). *Emergent Science: Teaching Science from Birth to 8*. London: Routledge.

Joyce, B. & Showers. B. (2002). *Designing Training and Peer Coaching: Our Needs for Learning*. Alexandria, VA: ASCD.

Kim, M. & Tan, A.-L. (2011). Rethinking difficulties of teaching inquiry-based practical work: Stories from elementary pre-service teachers. *International Journal of Science Education*, *33*(4), 465-486.

Kirschner, P. A., Sweller, J. & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, *41*(2), 75-86.

Klahr, D. (2009). "To every thing there is a season, and a time to every purpose under the heavens": What about direct instruction? In Tobias, S. & Duffy, T. M. (Eds), *Constructivist Instruction* (pp. 303-322). New York: Routledge.

Knight, S. (2009). *Forest Schools and Outdoor Learning in the Early Years*. London: SAGE.

Kruit, P. M., Oostdam, R. J., van den Berg, E. & Schuitema, J. A. (2018). Effects of explicit instruction on the acquisition of students' science inquiry skills in grades 5 and 6 of primary education. *International Journal of Science Education*, *40*(4), 421-441.

Kwan, R. & Wong, B. T. M. (2021). Latest advances in STEAM education research and practice: A review of the literature. *International Journal of Innovation and Learning*, *29*(3), 323-339.

Lindell, A. K. & Kidd, E. (2011). Why right-brain teaching is half-witted: A critique of the misapplication of neuroscience to education. *Mind, Brain, and Education*, *5*(3), 121-127.

Loucks-Horsley, S., Stiles, K., Mundry, S., Love, N. & Hewson, P. (2010). *Designing Professional Development for Teachers of Science and Mathematics*. Thousand Oaks, CA: Corwin Press.

Louv, R. (2005). *Last Child in the Woods: Saving our Children from Nature-deficit Disorder.* Chapel Hill, NC: Algonquin Books of Chapel Hill.

Mant, J., Wilson, H. & Coates, D. (2007). The effect of increasing conceptual challenge in primary science lessons on pupils' achievement and engagement. *International Journal of Science Education*, *29*(14), 1707-1719.

Matthews, M. R. (2002). Constructivism and science education: A further appraisal. *Journal of Science Education and Technology*, *11*, 121-134.

Matthews, M. R. (2021). *History, Philosophy and Science Teaching: A Personal Story*. Singapore: Springer Nature.

Mayer, R. E. (2005). Cognitive theory of multimedia learning. In: Mayer, R. E. (Ed.), *The Cambridge Handbook of Multimedia Learning*, pp. 31-48. Cambridge: Cambridge University Press.

Merton, R. K. (1973). *The Sociology of Science: Theoretical and Empirical Investigations*. Chicago: University of Chicago Press.

Mohd Radzi, R., Abdullah, M. N. S. & Muruthi, K. (2017). Inquiry-discovery teaching approach as a means to remediate primary students' misconceptions about the phases of the Moon. In: Karpudewan, M., Md Zain, A. & Chandrasegaran, A. (Eds), *Overcoming Students' Misconceptions in Science*, pp. 71-87. Singapore: Springer.

Millar, R. (2011). Practical work. In J. Osborne & J. Dillon (Eds), *Good Practice in Science Teaching: What Research has to Say*, pp. 108-134. Maidenhead: Open University Press.

National Academies of Sciences, Engineering, and Medicine (2022). *Science and Engineering in Preschool Through Elementary Grades: The Brilliance of Children and the Strengths of Educators*. Washington, DC: National Academies Press.

Nuffield Foundation (2013). *Evaluation of the Science Bursaries for Schools and Colleges Programme: Final Report.* Unpublished internal document.

Ødegaard, M. (2003). Dramatic science. A critical review of drama in science education. *Studies in Science Education*, *39*(1), 75-101.

Ofsted (2021). *Research Review Series: Science.* https://www.gov.uk/government/publications/research-review-series-science/research-review-series-science.

Osborne, J. F. (1996). Beyond constructivism. *Science Education*, *80*(1), 53-82.

Passy, R. (2014). School gardens: Teaching and learning outside the front door. *Education 3-13*, *42*(1), 23-38,

Perkel, J. M. (2019). Paper lets scientists play with each other's results. *Nature*, *567*(7 March), 17.

Perry, T., Lea, R., Jørgensen, C. R., Cordingley, P., Shapiro, K. & Youdell, D. (2021). *Cognitive Science in the Classroom*. London: Education Endowment Foundation (EEF). https://educationendowmentfoundation.org.uk/evidence-summaries/evidence-reviews/cognitive-science-approaches-in-the-classroom/.

Pyle, A. & Danniels, E. (2017). A continuum of play-based learning: The role of the teacher in play-based pedagogy and the fear of hijacking play. *Early Education and Development*, *28*(3), 274-289.
Reid, N. (2009). Working memory and science education: Conclusions and implications. *Research in Science & Technological Education*, *27*(2), 245-250.
Reiss, M. J. (1993). *Science Education for a Pluralist Society*. Milton Keynes: Open University Press.
Reiss, M. J. (in press). The nature of science. In: *Learning to Teach Science in the Secondary School: A Companion to School Experience, 5th edn*, Hetherington, L. (Ed.). Abingdon: Routledge.
Reys, R. E. (2001). Curricular controversy in the math wars: A battle without winners. *Phi Delta Kappan*, *83*(3), 255-258.
Riedinger, K., Marbach-Ad, G., Randy McGinnis, J., Hestness, E. & Pease, R. (2011). Transforming elementary science teacher education by bridging formal and informal science education in an innovative science methods course. *Journal of Science Education and Technology*, *20*, 51-64.
Rivera Maulucci, M. S., Brown, B. A., Grey, S. T. & Sullivan, S. (2014). Urban middle school students' reflections on authentic science inquiry. *Journal of Research in Science Teaching, 51*(9), 1119-1149.
Ross, M. E. (1997). Scientists at play. *Science and Children*, *34*(8), 35-38.
Roth, W.-M. & Bowen, G. M. (2001). Professionals read graphs: A semiotic analysis. *Journal for Research in Mathematics Education*, *32*(2), 159-194.
Royal Society (2020). *History of the Royal Society*. https://royalsociety.org/about-us/history/.
Rumelhart, D. E. & Ortony, A. (1977). The representation of knowledge in memory. In Anderson, R. C., Spiro, R. J. & Montague, W. E. (Eds), *Schooling and the Acquisition of Knowledge*, pp. 99-135. Hillsdale, NJ: Lawrence Erlbaum.
Russell, T. & McGuigan, L. (2016). *Exploring Science with Young Children: A Developmental Perspective*. London: SAGE.
Saylan, A., Armagan, F. Ö. & Bektas, O. (2016). The relationship between pre-service science teachers' epistemological beliefs and preferences for creating a constructivist learning environment. *European Journal of Science and Mathematics Education*, *4*(2), 251-267.
Sheldrake, R., Amos, R. & Reiss, M. J. (2019) *Children and Nature: A Research Evaluation for the Wildlife Trusts.* Newark: The Wildlife Trusts. https://www.wildlifetrusts.org/sites/default/files/2019-11/Children%20and%20Nature%20-%20UCL%20and%20The%20Wildlife%20Trusts%20Full%20Report.pdf.
Sikes, S. S. & Schwartz-Bloom, R. D. (2009). Direction discovery: A science enrichment program for high school students. *Biochemistry and Molecular Biology Education*, *37*(2), 77-83.
Sonnert, G., Michaels, M. & Sadler, P. (2013). Gender aspects of participation, support, and success in a state science fair. *School Science and Mathematics*, *113*(3), 135-143.
Stamp, K. & Preston, C. (Eds) (2021). *Teaching Primary Science Constructively, 7th edn*. South Melbourne: Cengage Learning Australia.
Stephenson, P. & Warwick, P. (2001). Peer tutoring in the primary science classroom: Upper primary students teaching infants. *Investigating*, *17*(2), 11-13.

Stylianidou, F. (2014). *Creativity in Science and Mathematics Education for Young Children: Executive Summary*. http://www.creative-little-scientists.eu/sites/default/files/Creativity_in_Science_and_Mathematics_Education.pdf.

Sutton, C. (1992). *Words, Science and Learning.* Milton Keynes: Open University Press.

Sutton, C. (1996). Beliefs about science and beliefs about language. *International Journal of Science Education, 18*(1), 1-18.

Tiberghien, A. (2000). Designing teaching situations in the secondary school. In Millar, R., Leach, J. & Osborne, J. (Eds), *Improving Science Education: The Contribution of Research*, pp. 27-47. Buckingham: Open University Press.

Tifi, A., Natale, N. & Lombardi, A. (2006). Scientists at play: Teaching science process skills. *Science in School*, *1*(2), 37-40.

Torgerson, C., Brooks, G., Gascoine, L. & Higgins, S. (2019). Phonics: Reading policy and the evidence of effectiveness from a systematic 'tertiary' review. *Research Papers in Education*, *34*(2), 208-238.

Tunnicliffe, S. D. (2013). *Talking and Doing Science in the Early Years: A Practical Guide for Ages 2-7*. Abingdon: Routledge.

Tunnicliffe, S. D. & Reiss, M. J. (1999). Building a model of the environment: How do children see animals? *Journal of Biological Education*, *33*, 142-148.

Tunnicliffe, S. D. & Reiss, M. J. (2000). Building a model of the environment: How do children see plants? *Journal of Biological Education*, *34*, 172-177.

Voon, X. P., Wong, L. H., Looi, C. K. & Chen, W. (2020). Constructivism-informed variation theory lesson designs in enriching and elevating science learning: Case studies of seamless learning design. *Journal of Research in Science Teaching*, *57*(10), 1531-1553.

Waite, S. (Ed.) (2011). *Children Learning Outside the Classroom: From Birth to Eleven*. Los Angeles: SAGE.

Weiss-Magasic, C. (2012). Writing and science literacy. *The Science Teacher*, *79*(1), 41-43.

Yasar, S. & Baker, D. (2003). Impact of involvement in a science fair on seventh grade students. Paper presented at the Annual Meeting of the National Association for Research in Science Teaching, Philadelphia, PA, USA.

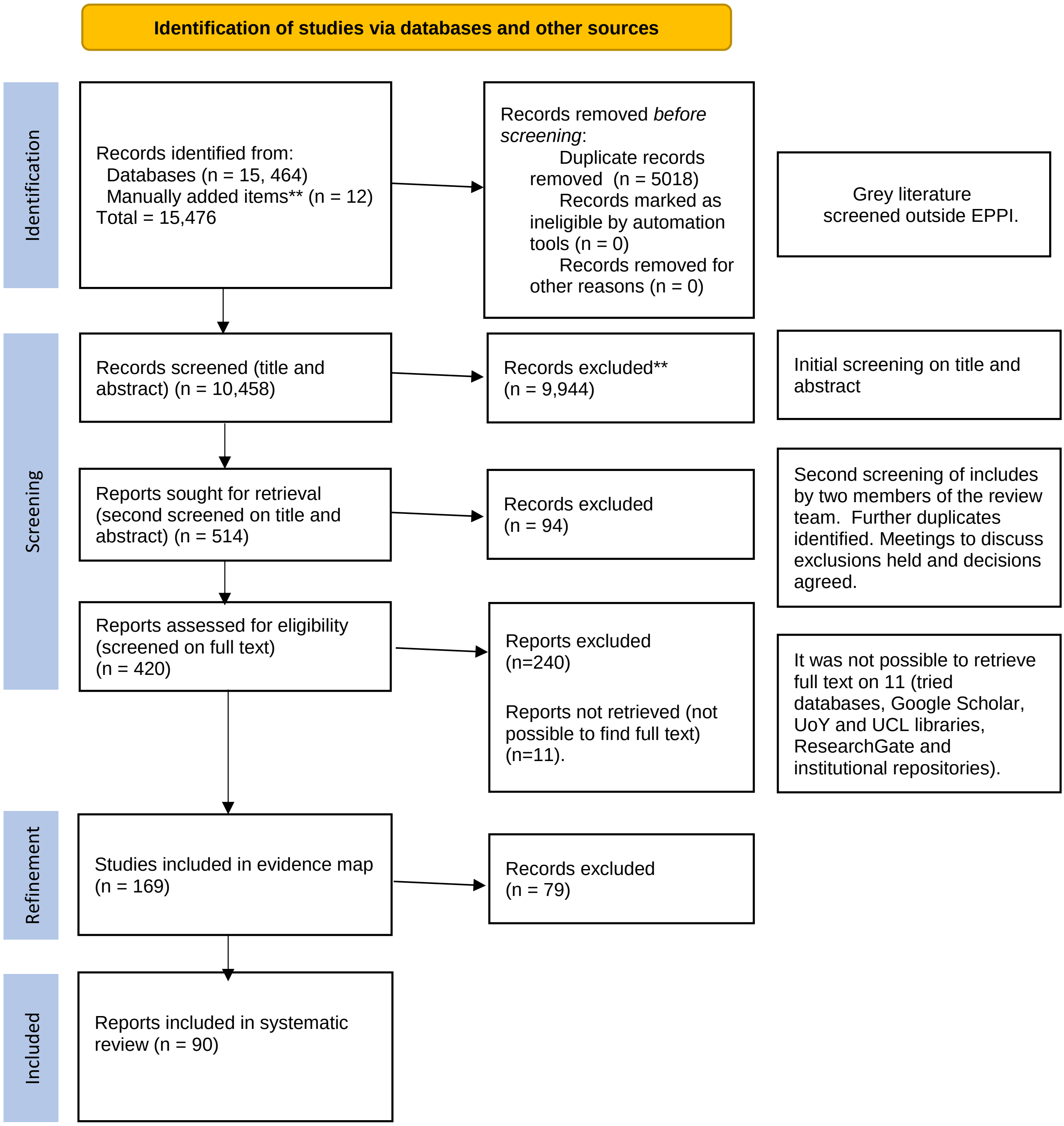
Young, M. F. D. (2008). *Bring Knowledge Back in: From Social Constructivism to Social Realism in the Sociology of Knowledge*. London: Routledge.

Young, M. (2014). Knowledge, curriculum and the future school. In Young, M. & Lambert, D. with Roberts, C. & Roberts, M. *Knowledge and the Future School: Curriculum and Social Justice* (pp. 9-40). London: Bloomsbury.

Young, T. (1994). I know what I want to teach them and I can't waste time doing what they want to do. *Education 3-13*, *22*(2), 22-25.

Appendix 4: PRISMA[2]

PRISMA Diagram



From: Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71. doi: 10.1136/bmj.n71. For more information, visit: http://www.prisma-statement.org/

[2] See example templates here: http://prisma-statement.org/PRISMAStatement/FlowDiagram.aspx

Appendix 5: Data extraction tool

Section 1: What is the publication type?
Journal article
Conference paper
Dissertation or theses
Technical report
Book or book chapter
Other (please specify)

Section 2: What is the research design and intervention approach?

What is/are the research question(s)? (Highlight)

What is/are the conceptual/theoretical underpinnings of the intervention? (Highlight)

How is the control group described? (e.g. BAU, enquiry approach) Highlight

Is there more than one treatment group?
Yes (please specify)
No
Not specified or N/A

How were participants allocated?
Random allocation (please specify)
Non-random but matched
Non-random, not matched prior to treatment
Not assigned – naturally occurring sample (Prospective QED)
Unclear

What was the level of allocation?
Individual
Class
School sample is in intervention OR control group
School is level of assignment and includes intervention and control groups
Region or district
Not provided/not available
Not applicable

What is the research method for this study?
Qualitative methods only
Quantitative methods only
Mixed methods
Other

What is the intervention aim/intended outcome? (Highlight)

Which approaches to teaching are identified? (Highlight detail)
Assessment
Feedback
Context-based approaches

Cross-curricular approaches
Learning outside the classroom
Co-operative or collaborative learning
Discussion
Questioning
Peer teaching
ICT supported teaching and learning
Language and literacy-based approaches
Text-based approaches
Mastery approaches
Multiple representations
Direct teaching
Practical work
Inquiry and investigation
Critical thinking and argumentation
Self-regulated learning approaches
Supplementary support
Specific named approaches
Other approach

In which cluster is/are the intervention approach(es)? (Select ALL that apply)
Assessment and feedback
Context based and cross-curricular approaches
Direct teaching, mastery, multiple representations and supplementary support
Practical work, inquiry and investigation and learning outside the classroom
ICT supported and online teaching and learning
Co-operative and collaborative approaches and peer teaching
Critical thinking and argumentation
Language, literacy and text-based approaches

What is the intervention name?

Section 3: Where did the study take place?

In which country/countries was the study carried out? (select ALL that apply)
All individual countries are listed from A-Z

Please add specific information about the location (highlight)
Specific to the location or place
Information about the type of location
No information provided

What is the educational setting (select ALL that apply)
Nursery school/pre-school/kindergarten
Infant/junior/primary/elementary school
Middle school
Residential/boarding school
Independent/private school
Home
Outdoor adventure setting
Other educational setting (please specify)

No information provided

Section 4: What is the relevant sample of the study?

What is the overall size of the relevant sample analysed?

What is the sex of the children?
Female only
Male only
Mixed sex
No information provided

What is the age of the children? (Select ALL that apply)
3-12 (listed separately)
No information provided

What is the proportion of children with low SES/FSM in the sample?
FSM or low SES student percentage
Further information about FSM or SES in the study sample
Related information in non-British context
No SES/FSM information provided

What are the characteristics of the study sample? (Highlight)
Code notes: Highlight any additional info to note about the characteristics of the sample. Examples include ethnicity/race, specific barriers to learning, additional learning needs, ELL

Are there any reported differences in characteristics between intervention and control groups? (Highlight)
Code notes: Does the study identify any characteristic differences between the control and treatment group(s)? e.g. sex, age, attainment scores

Section 5: What was involved in the intervention?

What type of organisation was responsible for providing the intervention?
School or group of schools
Charity or voluntary organisation
University/researcher
Local education authority or district
Private or commercial company
Other (please provide details)

Was training for the intervention/meeting with researchers provided?
Yes (please specify)
No
Unclear/Not specified

Who is the focus of the intervention? (Select ALL that apply)
Children
Teachers
Teaching assistants
Other educational practitioners

Non-teaching staff
Senior management
Parents
Other (please specify)

How are children organised? (Select ALL that apply)
Large group/class teaching (+6)
Small group/intensive support (3-5)
Paired learning
One to one
Child alone (self-administered)
Other (Explain in notes)

Were any of the following involved in the intervention or approach?
Digital technology
Yes
No

Parents or community volunteers
Yes
No

When did the intervention take place? (Select ALL that apply)
During regular school hours
Before/after school
Evenings and/or weekends
Summer/holiday period
Other (please specify)
Unclear/not specified

Who was responsible for the intervention at the point of delivery?
Teacher
Other (researcher/visitor)
Unclear/not specified

What was the role of the teacher at the point of delivery? (Highlight text)
Code notes: (Highlight text) Examples include: Generating questions; Setting up an enquiry/task; Facilitating discussion; Direct teaching; Assessing learning; Giving feedback

Were any teacher characteristics identified? (e.g. science lead, no. of years teaching) (Highlight)

What was the duration of the intervention? (Highlight)

What was the frequency of the intervention? (Highlight)

What was the length of the intervention sessions? (Highlight)

Are implementation details and/or fidelity details provided?
Qualitative
Quantitative
No implementation details provided

Are any costs for the intervention reported?
Yes (please add details)
No

Who undertook the outcomes evaluation?
The developer/researcher
A different organisation paid by developer
An organisation commissioned independently to evaluate
Unclear/not stated

Is this an EEF evaluation?
Yes
No
Section 6: Subjects tested and reported outcomes

What kind of tests were used? (Select ALL that apply).
Standardised data collection instrument (please add details, including if written, oral, IT-based, observational, interview)
Researcher developed data collection instrument (please add details, including if written, oral, IT-based, observational, interview)
School developed data collection instrument (please add details, including if written, oral, IT-based, observational, interview)
National test or examination (please specify)
International test (please specify)

Curriculum subjects tested (Select ALL that apply)
Science – highlight the topic (e.g. floating and sinking)
Science and English
Science and mathematics
STEM
STEAM
Other curriculum test

What were the outcomes of the teaching/intervention? Highlight detail
Conceptual understanding
Attainment
Recall
Progression
Thinking
Metacognition
Questioning
Confidence and/or self-efficacy
Attitudes
Motivation
Interest
Participation in or access to science
Relevance
Communication
Scientific language
Practical skills

Working scientifically
Teamwork
Creativity

In which cluster is/are the intervention outcome(s)? (Select ALL that apply) Exact clusters TBC

Conceptual understanding

Attainment, recall, progression

Thinking, metacognition, questioning

Confidence and/or self-efficacy
Attitudes
Motivation
Interest
Participation in or access to science
Relevance

Communication
Scientific language

Practical skills
Working scientifically
Teamwork
Creativity

What are the qual/quant analysis methods used? (Highlight)
Code notes: any statistical methods used and inductive/deductive qualitative methods

What are the qual/quant results from the study? (Highlight)
Code notes: Highlight p-values and effect sizes and/or key qual results

What are the key findings/outcomes from the study? (Highlight)

Section 7 Moderating factors

What factors are moderating the effectiveness of the intervention?

Barriers and facilitators linked to:
Learners, teachers, schools, system
(Highlight text)

Not mentioned (tick)

Appendix 6: Quality Assessment Tool

A consistent approach to quality assessment was used, drawing on the EHPP quality assessment tool. The following questions were to assess quality.

A) Selection bias

- (Q1) Are the individuals selected to participate in the study likely to be representative of the target population?
- (Q2) What percentage of selected individuals agreed to participate?
- Section Rating: Strong/Moderate/Weak

B) Study design

- Indicate the study design
- Was the study described as randomized? If NO, go to Component C CONFOUNDERS
- If Yes, was the method of randomization described?
- If Yes, was the method appropriate?
- Section Rating: Strong/Moderate/Weak

C) Confounders

- (Q1) Were there important differences between groups prior to the intervention? (hover over this cell to see examples)
- (Q2) **If yes**, indicate the percentage of relevant confounders that were controlled (either in the design (e.g. stratification, matching) or analysis)?
- Section Rating: Strong/Moderate/Weak

D) Blinding

- (Q1) Was (were) the outcome assessor(s) aware of the intervention or exposure status of participants?
- (Q2) Were the study participants aware of the research question?
- Section Rating: Strong/Moderate/Weak

E) Data collection methods

- (Q1) Were data collection tools shown to be valid?
- (Q2) Were data collection tools shown to be reliable?
- Section Rating: Strong/Moderate/Weak

F) Withdrawals and drop-outs

- (Q1) Were withdrawals and drop-outs reported in terms of numbers and/or reasons per group?
- (Q2) Indicate the percentage of participants completing the study. (If the percentage differs by groups, record the lowest).
- Section Rating: Strong/Moderate/Weak

Global Rating: Strong/Moderate/Weak

Studies were assessed to be strong if there were no weak ratings, moderate if there was one weak rating, and weak if there were two or more weak section ratings. Note: in the review text, strong corresponds to high quality, moderate to moderate quality and weak to low quality

Appendix 7: Summaries of study characteristics

EG: experimental group
CG: control group

N: number of pupils
RCT: randomised control trial

Funding and costs listed where reported

B1 Assessment and feedback in primary science teaching

Authors	Population, Setting and Context	Description of approach	Reported outcomes and measures
Decristan *et al.* (2015a)	Germany Funded (Hessian initiative for the development of scientific and economic excellence, LOEWE) Design: cluster RCT Age: 8-9 N: 551 (EG=319; CG=232) Topic: floating and sinking	Experimental group: Embedded formative assessments used at four points during the unit. Teacher set diagnostic tasks, gave written feedback on assessment, detailed pupil learning steps and differentiated tasks for pupils. Intervention at class level. Included teacher professional development (4 x 4.5 hours). Control group: Teaching as usual using the same unit as the experimental group (an inquiry-based unit on floating and sinking of objects lasting 5 lessons). Teachers were provided with resources including lesson plans and a manual.	Focus: attainment (knowledge and conceptual understanding) Measures of attainment gathered through previously developed instruments and a science competence test based on TIMSS 2007. Average level of science understanding of pupils in the experimental group was higher than that of the control group. Moderating effects of classroom process quality (cognitive activation, supportive climate, classroom management) on pupils' science understanding reported.
Decristan *et al.* (2015b)	Germany Design: Cluster RCT Age: 8-9 N: 873 (EG1=289; EG2=319; EG3=280; CG=232) Topic: floating and sinking	Three experimental groups: EG1: Scaffolding Instructional Discourse (SID) through talk EG2: Formative Assessment by task design and feedback EG3: Peer Assisted Learning (PAL) by forming dyads. Aimed to examine the effects of additional guidance, particularly for pupils with low language proficiency. Intervention at classroom level. Included teacher professional development (4 x 4.5 hours). Control group: Teaching as usual of an inquiry-based unit on floating sinking.	Focus: attainment (knowledge and conceptual understanding) Measures of conceptual understanding gathered through previously developed instruments and science attainment data from a test based on TIMSS 2007. Conceptual understanding improved in each intervention condition. Pupils in EG2 had significant gains in test performance compared to control group. No significant improvement seen in EG1 or EG3. Analysis suggests language proficiency is a moderating factor in the development of conceptual understanding, and that formative

Authors	Population, Setting and Context	Description of approach	Reported outcomes and measures
			assessment and scaffolding instructional discourse were beneficial for pupils with low language proficiency.
Ferrell *et al.* (2017)	USA Design: quasi-experiment Age: 8-13 N: 295 Topic: various	Experimental group: An audible image description provided in assessments for pupils with visual and print disabilities. Control group: No audible description provided. Intervention at individual level. 178 pupils with print disabilities, 117 with visual disabilities.	Focus: attainment Audible image description enhances attainment for pupils with visual disabilities who use braille (p=0.03), with a reported effect size=0.66. Audible image description does not enhance attainment for pupils with print disabilities. Audible image description does not enhance attainment for pupils with visual disabilities who use print (either regular print or large print).
Hondrich *et al*. (2018)	Germany Design: RCT Age: 8-9 N: 551 (EG=319; CG=232) Topic: floating and sinking Note links to Decristan *et al.* (2015)	Experimental group: Teachers given an adapted version of an inquiry unit. Adaptations included embedded formative assessment materials (a written task, semi-standardised individual feedback and adapted worksheets) for Unit 1 of two units. For Unit 2, they had to develop the formative assessment themselves. Additional 3 professional development workshops on formative assessment. Control group: same inquiry unit, but no sample formative assessment materials provided. Additional 3 CPD workshops on parental counselling. All teachers received standardised materials and a curriculum manual and 4 x 4.5 hours professional development.	Focus: attitudes, motivation Perceived competence and intrinsic motivation measured using questionnaires, assessed by scales adapted from Blumberg (2008) pre, post Unit 1 and post Unit 2. Pupils in the EG showed significant positive effects in intrinsic motivation. Post Unit 1, the difference between the two groups was at most marginally significant (p=0.07). Post Unit 2, the difference was more significant (p=0.03). Pupils in the EG showed significant positive effects in perceived confidence: post Unit 1 (p<0.05) and post Unit 2 (p<0.01). Higher perceived competence at post Unit 1 appeared to mediate the effect of formative assessment on intrinsic motivation at post Unit 2.
Hwang *et al.*(2018)	Taiwan Funded (Ministry of Science and Technology, Taiwan) Design: quasi-experiment	Experimental group: guided peer feedback in an e-book development activity. Pupils evaluated their peers' work using criteria provided. Control group: participated in an e-book development activity. Pupils received feedback from their teacher.	Focus: Attainment and critical thinking Measures of plant knowledge (researcher developed), innovative thinking (modified from Lin and Wang, 1994) and cognitive load (modified from Paas, 1992).

Authors	Population, Setting and Context	Description of approach	Reported outcomes and measures
	Age: 10 N: 72 (EG=36; CG=36) Topic: plants and ecology	Both groups created an e-book on a website.	Experimental group performed significantly better than control group on achievement (F=14.82, $p<0.001$), design (t=2.84, $p<0.001$), and content of e-books (t=6.51, $p<0.001$), innovative thinking (t=2.21, $p<0.05$). Cognitive load measures indicate that the intervention did not pose a substantially increased demand on pupils (t=-2.98, $p<0.01$).
Hwang *et al.* (2021)	Taiwan Funded (Ministry of Science and Technology, Taiwan) Design: quasi-experiment Age: 10 N: 101 (EG=52; CG=49) Topic: geology	Experimental group: bidirectional peer assessment. Pupils respond to peer feedback using online concept maps. Portable devices were used to produce the concept maps. Implemented over 9 sessions x 40 minutes over 2 weeks. Control group: conventional peer assessment with concept maps, without the chance to respond. Intervention at classroom level with the same teacher. Pupils' concept maps scored by two experienced natural science teachers. Note links to ICT and online learning.	Focus: attainment, attitudes, critical thinking Measuring tools included a science learning achievement test, learning motivation questionnaire (Wang & Chen, 2010), self-efficacy questionnaire adapted from Pintrich, Smith, Garcia, and McKeachie (1991), critical thinking tendency questionnaire adapted from Chai, Deng, Tsai, Koh, and Tsai (2015), environmental identity questionnaire adapted from Clayton (2003) and cognitive load questionnaire adapted from Hwang, Yang, and Wang (2013). Concept maps scored using a rubric modified from Saxton, Belanger, and Becker (2012). Positive effects of bidirectional peer assessment on learning achievement (F=34.31, $p<0.001$), concept mapping scores (F=4.02, $p<0.05$), self-efficacy (F=10.42, $p<0.01$), critical thinking tendency (F=22.56, $p<0.001$) and environmental identity (F=9.32, $p<0.01$) found. No significant differences between experimental and control group found on the cognitive load test, suggesting that the demand of bi-directional peer assessment was appropriate.

B2 Context-based and cross-curricular approaches to primary science teaching

Authors	Population, Setting and Context	Description of approach	Reported outcomes
Burt *et al.* (2022)	USA Design: quasi-experiment Age: 9-10 N: 3,121 (EG=638; CG1=993; CG2=1,490) Topic: climate change	Experimental group: the New York Sun Works (NYSW) Programme. Hydroponic gardening incorporated into all relevant topics, included an adapted classroom/lab with hydroponic/aquaponic technology. Teacher training provided (36 hours). Comparison groups: CG1 received the intervention curriculum the following year; pupils in CG2 attended matched schools. It is not clear what the comparison groups did in place of the NYSW curriculum.	Focus: attainment Measures: The New York State science achievement tests. Experimental group pupils scored significantly higher than matched control group pupils ($t=-10.93$, $p<0.001$), and than pupils in the peer group (CG2, $t=-11.73$, $p<0.001$) on the fourth-grade (age 10) science achievement test.
Fasasi (2017)	Nigeria Design: quasi-experiment Age: 9-12 N:352 Topics: You and Environment, Living and Non-Living Organisms, You and Science and You and Energy	Experimental group: ethnoscience instruction linking science education to culture. Includes phases of identification of cultural beliefs, sayings and practices relevant to the topic, and classification of these as compatible, modifiable or contradictory to science. Control group: modified lecture method, drawn from the basic science curriculum module of the Federal Ministry of Education. Includes teacher demonstrations. Teachers in both groups provided with a guide.	Focus: attitudes to science Attitude Towards Science Scale adapted from standardised Modified Sherman Science Attitude Scale (Doepken, Lawsky & Padwa, 2003) used to measure pupils' attitudes. Ethnoscience instruction improved attitudes towards science compared to the control group ($F(1,347)=296$, $p<0.5$). Effect size $\eta^2=0.46$. Impact of ethnoscience instruction on attitudes particularly positive for children in rural schools and those with parents with low educational status
Hardiman *et al.* (2017)	USA Funded (U.S. Department of Education) Design: RCT Age: 10-11 N: 350	Experimental group: arts-integrated science instruction unit involving demonstration of knowledge through visual and performing arts. Pupils reinforced knowledge and displayed understanding with use of arts such as rap, dance and drawing. Delivered over 15 days. Control group: conventional science instruction unit. Arts-integrated units were created, along with conventional units, matched for content, dosage and type of instructional	Focus: attainment Measures included multiple choice tests of curriculum content. No significant difference in retained content between the arts-integrated and conventional science groups ($F91,508)=0.128$, $p=0.72$). Pupils who were basic readers remembered more science content learnt in the arts-integrated method than conventional methods.

Authors	Population, Setting and Context	Description of approach	Reported outcomes
	Topic: astronomy, life science, chemistry and environmental science	activity. Teacher professional development (10 hours) provided. Each unit was taught over 3-4 weeks, with the experimental and control group experiencing both units (in reverse order).	Order appears to matter: students who experienced arts-integrated instruction first remembered more science in the second unit when they learned science using the conventional method.
Olgun & Adali (2008)	Turkey Design: quasi-experiment Age: 9-10 N: 88 (EG=43; CG=45) Topic: viruses, bacteria, fungi and protista	Experimental group: described as a 'case study approach'. An ill-structured problem related to viruses and bacteria was provided using a real-life scenario with conflicting information. Pupils were asked to solve the 'problem' by searching the library or internet prior to class. Discussion in class groups facilitated by the teacher. Teacher as 'metacognitive coach' encouraging pupils' contributions in class. Pupils write responses to follow-up questions individually. Professional development provided. Control group: reading assignment prior to lesson with one pupil reporting key constructs, then teacher explains and asks questions	Focus: attainment and attitude Attainment measured using a science achievement test pre- and post-intervention. Attitude measured using an attitude scale towards science. Pupils in the experimental group scored significantly higher than pupils in the control group in the test of achievement ($p<0.05$) and in attitudes towards school science ($p<0.05$).
Qiao & Zhou (2020)	China Funding (Ministry of Education Humanities and Social Sciences) Design: quasi-experiment Age: 10-11 N: 200 (EG=100; CG=100) Topic: buoyant force	Experimental group: STEM teaching method delivered over 8 hours. Unit integrates science, technology, engineering and mathematics. Intervention at classroom level. Control group: 'traditional science teaching'. No course content provided for the control group.	Focus: attainment Measures of attainment collected using questionnaire/test developed by the research team. Experimental pupils showed significantly better results than control group pupils in total score, basic knowledge and ability expansion ($p<0.01$).

Authors	Population, Setting and Context	Description of approach	Reported outcomes
Zhang & Campbell (2012)	China Design: quasi-experiment Age: 8-11 N: 385 (EG=201; CG=184) Topic: not stated	Experimental group: Integrated Experiential Learning Curriculum (IELC), employing experiential learning emphasising the importance of pupils applying school learning to authentic problem-solving tasks. Professional development (50 hours total) over 2 weeks prior to implementation. One year intervention period. Control group had normal teaching (business as usual). Intervention at classroom level.	Focus: attitude Measures of attitude were collected using the three-dimension elementary science attitude survey (Zhang & Campbell, 2011) which assesses cognitive, affective and behavioural components of attitudes. Citizen beliefs also measured pre- and post-intervention. The experimental group showed greater improvement than the control group in attitudes towards science and the learning environment ($p<0.005$). The study also suggests that the IELC treatment had a positive effect on pupils' citizenship beliefs and attitudes towards the learning environment.

B3 Co-operative and collaborative approaches

Authors	Population, Setting and Context	Description of approach	Reported outcomes
Chang & Hsin (2021)	Taiwan Design: quasi-experiment Age: 10-11 N=104 (EG=52; CG=52) Topic: position and motion of the sun	Experimental group: Worksheets followed by distinct phases of self-explain, discuss and re-explain designed to help pupils learn science and foster the habit of independent thinking. Control group: traditional teacher-centred lecture approach. They used the same worksheets as the EG but in the CG, the teacher allowed pupils to ask questions about what they did not understand, which she then answered.	Focus: attainment Two-tier multiple-choice test developed on the basis of the Teacher's Manuals for Elementary School Science and Technology used pre- and post-intervention. The low-achievers in the EG had significantly higher post-test scores than in the control group. The low-achievers in the experimental group had significantly higher post-test scores (reported effect size 2.04) than those in the control group (reported effect size 1.4). The high-achievers in the experimental group had significantly higher post-test scores (reported effect size 2.27) than the high-achievers in the CG (reported effect size 1.58). Both groups made significant gains but there was a ceiling effect for high-achievers. This indicates that the SDR strategy was effective in helping close the achievement gap between high- and low-achievers.
Chen *et al.* (2013)	USA Design: quasi-experiment Age: 9-10 N: 454 (EG=145; CG=309) Topic: forces and motion	Experimental group: letter exchange between pupils aged 9 and 16/17 (Write to Learn project). Letters focus on 'argumentative writing' – constructing persuasive arguments by using questions, claims, data, and evidence. Eight-week intervention. Pupils in groups of two to four. Included teacher professional development. Control group: standard curriculum (business as usual). Intervention delivered at classroom level. Evaluation of written letters by external evaluators.	Focus: attainment Measures include pre and post multiple-choice questions developed by Horizon Research Institution (developer of tests for use in schools). Pupils in the experimental group performed statistically significantly better on multiple-choice questions than those in the control group (effect size=0.25). Analysis of sub-groups showed significant differences, with effect sizes for socio-economic status (0.41) and gender (0.35).

Authors	Population, Setting and Context	Description of approach	Reported outcomes
Eysink *et al.* (2017)	The Netherlands Funded (Netherlands Organisation for Educational Research) Design: quasi-experiment Age: 8-11 N: 306 Topics: (a) magnetism, (b) Sun, Earth, and Moon, (c) buoyancy, (d) sound, (e) senses, (f) weather	Experimental group: differentiation (STIP approach, translated as Collaboration during differentiation in Task, Content, and Process). Involves setting work according to teachers' ranking (based on scores on Dutch student monitoring system, as well as teacher judgement), with the lowest 20%, highest 20% and remaining 60% constituting the three groups. Pupils work in a mixture of homogeneous and heterogeneous groups on tasks with a common group goal. Teacher assumes role of coach. Control group: Regular instructional approach, usually textbook with text and illustrations and exercise book with assignments without intervention. Intervention at classroom level, size of pupil groups not given.	Focus: attainment Researcher-developed tests used to measure pupils' domain knowledge and the instructional value pupils placed on STIP approach. Initial statistical analysis of pupils' domain knowledge indicated that the experimental group performed statistically significantly *worse* than the control group ($t(211.94)=3.40$, $p<0.001$, $d=0.46$). This finding was reversed when the analysis was limited to the teachers who had scored highly on differentiation activity based on classroom observations. Note: the control group teachers were more experienced than the experimental group. Experimental group pupils appeared to have a less positive attitude to the STIP approach, initially, but this improved with time.
Hand *et al.* (2018)	USA Design: cluster RCT Age: 7-10 N: 9963 (for attainment tests) N: 2353 (for critical thinking tests) No specific numbers given for EG and CG, but 24 schools in each of EG and CG. Topic: various	Experimental group: Instruction involving teaching pupils to use science argument as a core component of building understanding of the content and developing critical thinking skills (Science Writing Heuristic-SWH). Two-year intervention, linked to the implementation of the Next Generation Science Standards (NGSS). Control group: Normal curriculum (business as usual), though paper states teachers received CPD specific to their district. Extensive teacher professional development. Intervention at classroom level.	Focus: attainment, critical thinking Impact on pupils measured via standard national tests: attainment measured via Iowa Test of Basic Skills (ITBS), Iowa Assessments Test (IAT); critical thinking measured by Cornell Critical Thinking (CCT) test. Science attainment: no significant gains in science scores for experimental group when compared with control group. Critical thinking: significant gains in pupils' critical thinking for experimental group when compared with control group ($p<0.05$, effect size=0.17). Effect sizes were more pronounced for disadvantaged pupils: Pupils working in an additional language: effect size=0.22. Free or reduced-price school meals: effect size=0.18. Pupils with individualized education plan: effect size=0.29.

Authors	Population, Setting and Context	Description of approach	Reported outcomes
Looi *et al.* (2010)	Singapore Funded (National Research Foundation, Singapore) Design: quasi-experiment Age: 10 N: 240 (EG=160; CG=80) Topics: circulatory system, energy, light, heat	Experimental group: collaborative activities incorporated the use of a computerised version of sticky notes (Group Scribbles – GS). Implemented over 10 weeks (1-hour traditional science lesson and 1-hour Group Scribbles lesson in a computer lab). Intervention at classroom level. Control group described as not participating in collaborative Group Scribbles activities.	Focus: attainment, attitudes Researcher-developed pre and post questionnaire about attitudes towards science learning, epistemology, and collaborative learning (and electronic sticky notes in post-questionnaire). Experimental group pupils in classes for pupils with high prior attainment performed significantly better in science attainment tests than control group classes for pupils with high prior attainment. The researchers report that the intervention brought positive changes to pupils' attitudes to learning science and to their ability to work collaboratively.
Reeves *et al.* (2013)	USA Design: cluster RCT Age: 9-10 N: 4,713 Topic: various	Experimental group: guided inquiry-based approach together with the development of argumentation skills via the Science Writing Heuristic (SWH) approach. Three days of CPD provided on how to foster argumentation and inquiry through co-operative and individual learning. Pupils work in groups. Control group: guided inquiry-based approach without science writing heuristic.	Focus: attainment Study explores impact on science achievement as measured by standardised Iowa Test of Basic Skills (ITBS). Findings reported as path coefficients, rather than effect sizes. All effects are reported to be significant at $p<0.01$ level. Results focus on pupils with differing backgrounds. Information on science attainment is limited – the figures quoted appear to be for achievement across English, mathematics and science, and the paper text suggests there was little impact on science attainment. Path coefficient for (i) all pupils=0.81, (ii) pupils with special educational needs=1.12 (with effects reported as becoming increasingly beneficial with increasing exposure), (ii) pupils labelled gifted and talented=0.72.

B4 Critical thinking and argumentation in primary science

Authors	Population, Setting and Context	Description of approach	Reported outcomes
Arias *et al.* (2017)	USA Funded (National Science Foundation) Design: Quasi-experiment Age: 8-11 N: 1,152 (EG=646; CG=506) Topics: electric circuits and ecosystems	Experimental group: teachers used materials (two commercially available kit-based inquiry-oriented units) with researcher-enhanced educative features including making predictions with justification. These focused on the integration of science practices and content. Teachers in the control group used the original curriculum materials: two commercially available kit-based inquiry-oriented units. Both groups of teachers participated in professional development (2 days).	Focus: critical thinking Measures included pre-and post-unit student assessments and samples of student work. Experimental group pupils demonstrated a significant increase in the number of predictions that included justifications and a significant increase in clear, aligned, and accurate justifications. The effect size for both units was 0.27. Pupils' notebooks and lesson videos suggested that the teachers in the treatment group were supporting pupils in making justified predictions in ways that teachers in the control group were not.
Chen *et al.* (2016)	Taiwan Funded (Ministry of Science and Technology) Design: quasi-experiment Age: 9-10 N: 72 (EG=36; CG=36) Topics: sound, magnetic force, capillarity, light, gravity, and static electricity	Experimental group: Modified Argument Driven Inquiry (ADI) sessions. This involved 12 x 100-minute sessions and: identifying a focus task from a demonstration or presentation, identifying related research questions, making hypotheses, designing an investigation and procedures, collecting data, providing evidence-based conclusions, sharing a group argument and critiquing and refining explanations. Intervention took place over 12 weeks. Control group: Regular science lessons. The lessons were teacher-directed considerations of the textbook supplemented with teacher presentations, completion of study guides, and occasional demonstration or cookbook practical work.	Focus: attitudes, critical thinking Student questionnaire included demographic data, engagement of learning science scale (ELSS) and an argumentation test. The study found that the adjusted post-test engagement of learning science scores of the experimental group students were significantly higher than the control group on the total engagement of learning science scale ($F(1, 69)=4.74$, $p=.033$, $\eta^2=0.06$). Adjusted post-test scores of the experimental group pupils were significantly higher than the control group on total argumentation score ($F(1, 69)=10.29$, $p=.002$, $\eta^2=0.13$). Note: the control group did not cover the same topics as the experimental group.

Authors	Population, Setting and Context	Description of approach	Reported outcomes
Kara & Kingir (2021)	Turkey Design: quasi-experiment Age: 9-10 N: 107 (EG=55; CG=52) Topic: properties of matter, lighting and sound technologies	Experimental group: used a model-based science writing heuristic. Pupils build arguments using models, inquiry and negotiations. Science writing heuristic introduced through a murder mystery game. Intervention at classroom level. Control group: traditional approach teaching the same units. Both groups studied science over 17 weeks (3 x 40-minute sessions/week). Pupils in groups of 4-5.	Focus: attainment, critical thinking Tests measuring students' conceptual understanding of properties of matter and lighting and sound technologies units were used in addition to a researcher-developed rubric for scoring pupils' written answers. Experimental group pupils performed significantly better than the control group on attainment tests for both topics (effect size=0.91, combined for both topics) Experimental group pupils' construction of question, model, claim, and evidence improved over the duration of the intervention. Study suggests that quality of writing seems to relate to conceptual understanding.
Miller *et al.* (2014)	USA Funding: Institute of Education Sciences, U.S. Department of Education Design: quasi-experiment Age: 8-10 N: 130 Topics: heat transfer and buoyancy	Experimental group: argumentative discussion in the context of reading a text (based on concept cartoons (Keogh & Naylor, 1999)). Teachers participated in a 3-hour professional development workshop. A collaborative reasoning discussion treatment was applied over two weeks (with two weeks of testing before and after the intervention). The comparison groups involved pupils reading for no stated purpose or to prepare for a regular classroom discussion.	Focus: attainment, thinking Measures included questions about a shape of the Earth and a scored interview. Pre-tests included the Stanford Achievement Test 10th Edition reading comprehension subtest and a CogAt (Riverside Publishing, 2001) figure analysis subtest (chosen to be analogous to imagining Earth from different perspectives). Pupils with consistent views in this study became more entrenched in their beliefs if they anticipated an argumentative discussion.

Authors	Population, Setting and Context	Description of approach	Reported outcomes
Sternberg *et al.* (2014)	USA Funded (National Science Foundation) Design: quasi-experiment Age: 9-10 N:7,702 Topics: nature of light, magnetism	Experimental group: units of instruction based upon the theory of successful intelligence (SI; analytical, creative, and practical instruction). Comparison groups included teaching as usual (weak control), memory instruction (strong control) and critical-thinking instruction (strong control). Same content was covered by all groups, but with different activities used. All teachers participated in 12 hours of CPD delivered over 2 days. Note: we focus here only on science, which is a small part of the overall study.	Focus: attainment Attainment was measured using unit-specific assessments graded using rubrics. For the Nature of Light unit, there was a significant intervention effect in favour of successful intelligence over the memory condition ($p<0.01$) and for Magnetism, there was a significant advantage for the critical thinking condition over SI ($p=0.04$), suggesting that whilst successful intelligence is not advantageous, nor is it disadvantageous for science learning. Controlling for accessible demographic characteristics of schools and classrooms, the SI intervention was found to be advantageous but weakly and inconsistently so.
Tsai *et al.* (2012)	Taiwan Funded (National Science Council) Design: quasi-experiment Age 10-11 N=189 (EG1=64; EG2=65; CG=60) Topic: vision	Experimental group: the intervention groups used a cognitive apprenticeship web-based argumentation (CAWA) system. This system instructed pupils on the Toulmin Argumentation Pattern. The intervention was delivered over 12 weeks. EG1: instruction included enhanced web tools. EG2: instruction on the CAWA system only. Control group: Arguments on paper. Linked to IT-supported and online-learning cluster.	Focus: critical thinking Measures of argumentation performance gathered using three researcher-designed argumentation tasks administered at baseline and post-treatment and at a delayed point (different tasks at each time point). Integration of argument, use of evidence and evaluation of argument measured. Pupils in EG1 demonstrated better performance than pupils in EG2 and pupils from both experimental groups demonstrated better performance than pupils in the control group. The explicit scaffolds and hints in the computer system appears to have helped pupils understand how to make an argument.

B5 Explicit instruction and related approaches in primary science

Study	Location and sample details	Description of approach	Reported outcomes
Baumfalk *et al.* (2019)	USA Funded (National Science Foundation) Design: quasi-experiment Age: 8-9 N: 201 (EG=85; CG=116) Topic: the water cycle	Experimental group: Model enhanced version of the Full Option Science System (FOSS) water unit. For each of 4 investigations, pupils participated in modelling practices (constructing, using, evaluating and revising models) and created model-based explanations. Control group: Full Option Science System water unit without modelling enhancements (structured worksheets). Study took place in 5 schools with 11 classes.	Focus: working scientifically Modelling task pre- and post-intervention, scored using rubric designed by researchers. The model-enhanced group scored significantly better than the comparison group on the combined dependent variables after controlling for pre-unit scores for epistemic features: components, sequences, mapping, explanatory process, and scientific principle ($F=7.911$, $p<0.001$, $\eta^2= 0.177$). Specifically, children in the intervention group emphasised non visible components of the water cycle (e.g. groundwater), identified greater numbers of sequences, and used components such as the sun to explain processes such as evaporation.
Berry *et al.*(2013)	USA Design: quasi-experiment Age: 8-9 N: 58 (EG=29; CG=29) Topic: soil formation	Experimental group: concept maps used as a pre-reading activity, shared with classmates, then the class constructed a concept map together. Second concept map created post-activity. Control group: teacher questioning as a pre-reading strategy, with pupils writing their answers down then sharing with classmates. Teacher posts answers on the board. Study took place in one school during an informational read-aloud activity. Performance on pre-test was comparable between both groups.	Focus: attainment Measures include researcher-developed tests of relational knowledge, vocabulary matching and comprehension and writing comprehension. The study found that the treatment group outperformed the comparison group on relational vocabulary, multiple choice (identification of key ideas) and written expression at levels of statistical significance ($p<0.001$).

Study	Location and sample details	Description of approach	Reported outcomes
Cohen & Johnson (2012)	USA Design: quasi-experiment Age: 10-11 N: 89 Topic: biology	Experimental groups: test of dual coding theory. Very brief intervention. Three conditions included: (i) a picture presentation intervention, pairing pictures with the vocabulary; (ii) an image creation – no picture intervention in which pupils were asked to imagine an image to go with the vocabulary and draw it on paper; and (iii) an image creation – picture intervention in which pupils were presented with pictures and asked to draw them on paper. Control group: the comparison group was given a simple verbal presentation of the scientific term and concepts. Note: The intervention appears to have taken place out of context, i.e. words not necessarily related to the science being taught as part of a science unit.	Focus: scientific language Measures included the Peabody Picture Vocabulary Test, a vocabulary test including three distractors and 'not sure/do not remember', a word fill-in task and a definition word match. The group assigned to the image creation conditions (both those provided with the image and those not) performed significantly better than the picture presentation and word repetition conditions, suggesting that child-created imagery facilitates learning of science vocabulary. Sample controlled for demographic variables in a largely monolingual sample.
Doabler *et al.* (2021)	USA Funded (National Science Foundation) Design: cluster RCT Age: 7-8 N: 291 (EG=141; CG=150) Topic: Earth science	Experimental group: 'Science Explorers' curriculum unit using explicit instruction (guided inquiry), specifically 1) activation of prior knowledge and context (2) explicit vocabulary instruction (3) read-alouds and (4) investigation activities. Implemented over 10 x ~40-minute lessons (5/week). Teachers participated in 6 hours of CPD before the start of the study. Control group: Teaching as usual (blend of district-developed materials and commercially available science programmes). Study took place in 3 schools.	Focus: attainment, working scientifically, scientific language Measures include content knowledge and scientific practices, a test of early geoscience learning, a virtual interactive science practices assessment and a measure of science vocabulary knowledge. Large effect on vocabulary test outcomes for the experimental group (g=0.94). Positive effects also on geoscience learning (g=0.60 and virtual interactive science practices (VISPA) (g=0.48). Non-significant effects on content knowledge and scientific practices. No moderation effects found.

Study	Location and sample details	Description of approach	Reported outcomes
Kim *et al.* (2012)	USA Design: quasi-experiment Age: 5-9 N: 2,182 (EG=1,224; CG=958) Topic: inquiring about change and systems.	Experimental group: higher-level inquiry-based curriculum emphasising "doing" science and deep learning through concept mastery and investigation. Units integrate higher-level questions, reflection and discussion. Control group: business as usual assumed. Study took place in 6 schools. Data presented from three years of implementation. Professional development at least twice a year, with project staff as ambassadors to support teachers during implementation.	Focus: attainment, critical thinking Measures included the Bracken Basic Concept Scale – Revised, the Naglieri Nonverbal Intelligence Test, the Metropolitan Achievement Test (8th edition, MAT-8) science subtest (standardised test) and the test of critical thinking (standardised test). No difference in science understanding between experimental and control groups in years 1 and 3; medium effect (Partial η^2=0.013) in year 2. No difference in critical thinking in year 1, but medium effect on experimental group in years 2 and 3 (Partial η^2=0.03). Positive effect reported for pupils from low socio-economic backgrounds.
Michalsky *et al.* (2009)	Israel Design: quasi-experiment Age: 9-10 N: 108 Topic: animals and plants	Metacognitive instruction before, during or after a cooperative group reading of a scientific text, three times a week for 12 weeks. The instruction consisted of questions on comprehension of the text, connections between existing and new knowledge, appropriate inquiry strategies to solve the problem and reflection on the solution. Comparison group: no metacognitive instruction. Study took place in 4 classes in 4 schools, all using the same inquiry curriculum and textbooks and involving pupils working in collaborative learning groups.	Focus: attainment, critical thinking, working scientifically Measures included a test of scientific knowledge designed by Ministry of Education, test based on PISA scientific literacy test, and an adapted version of the Metacognition Awareness Questionnaire (Schraw & Dennison (1994). There was a significant difference between participants placed in the different intervention groups at immediate ($F(3,84)=6.94$, $p<0.001$) and delayed recall. The metacognitive conditions were found to be more effective than instruction without metacognitive intervention across all measures. Metacognitive instruction after reading the scientific text significantly outperformed all groups; metacognitive instruction before reading the texts outperformed metacognitive instruction during reading, and the no metacognitive instruction obtained the lowest scores.

Study	Location and sample details	Description of approach	Reported outcomes
Randler (2009)	Germany Design: quasi-experiment Age: 6-10 N: 138 (EG=67; CG=71) Topic: bird identification	Experimental group: Pupils were given an identification book and worked with taxidermic specimens at workstations to complete worksheets. Comparison group: As in the experimental group, but with soft toys rather than taxidermic specimens. No differences between experimental and control groups at pre-test.	Focus: attainment The measure used was a test to assess children's ability to identify six bird species from photographs, with different tests used prior to the intervention, post-intervention and on the retention test. Worksheets also analysed. No differences between groups in the post-test, retention test, or on the worksheets, suggesting soft toys are as effective as taxidermic specimens. Some differences found in favour of taxidermic specimens using a general linear model, but these had disappeared by the retention test.
Rotgans & Schmidt (2017)	Singapore Funded (National Institute of Education (Nanyang Technological University) Singapore) Design: quasi-experiment Age: 9-10 N: 129 Topic: properties of light	Experimental group: pupils were given situational interest-inducing science problems and asked to set the learning goals in each of 2 sessions a week over four weeks. Control group: the comparison group was provided with similar information by the teacher. The study took place in 4 classes in 1 school over 4 weeks. There were no significant differences between groups in relation to sex, age or prior knowledge.	Focus: attitudes Measures included the individual interest questionnaire (Rotgans, 2015) and situational interest questionnaire (Rotgans & Schmidt, 2011; 2014). Results analysed using latent growth curve modelling. The study found that students in the group receiving the problems experienced a growth in individual interest over the 4 weeks, whereas students in the control group lost interest.

Study	Location and sample details	Description of approach	Reported outcomes
Upadhyay & DeFranco (2008)	USA Funded (DNR Minnesota) Design: quasi-experiment Age: 8-9 N: 108 (EG=51; CG=57) Topic: environmental science	Experimental group: connected instruction (design and implementation of science instructions that connect children's funds of knowledge), supported by professional development with researchers. Comparison group: direct instruction (not defined) compared with connected science instruction Study conducted in 4 classes comparable in size, demographic features and teacher experience.	Focus: scientific language Measures include a researcher-designed environmental science knowledge and vocabulary test. The studies found that pupils in the connected science instruction group showed lower gains in environmental science vocabulary than the direct instruction group, and that the loss of vocabulary was greater in the direct instruction group.
van der Graaf *et al.* (2019)	Netherlands Funded (NRO) Design: randomised control trial Age: 9-10 N: 301 Topic: control of variables; construction	Experimental group: Direct instruction combined with verbal support teacher training. Implemented in the context of scientific reasoning during a series of inquiry-based lessons on the control of variables strategy on solidity of constructions. The direct instruction addressed hypothesis construction and evidence evaluation. Comparison group: direct instruction only, verbal support only, lesson-series (inquiry) only (baseline condition). All conditions were enacted in an inquiry-based programme of 6 lessons lasting about 60 minutes each. Teachers were trained in verbal support during a 3 hour session.	Focus:attainment, critical thinking, scientific language Measures included a scientific reasoning inventory adapted by researchers, an academic and domain-specific vocabulary test and a researcher-designed domain specific knowledge test (near and far transfer). These were applied pre- and post-intervention, and at a delayed time point. The study found a positive effect of direct instruction on scientific reasoning and verbal support during inquiry improved the effectiveness of an inquiry. Combining direct instruction and verbal support had the largest effect.

Study	Location and sample details	Description of approach	Reported outcomes
Williams *et al.* (2009)	USA Funded (U.S. Department of Education, Office of Special Education Programs) Design: quasi-experiment Age: 7-8 N: 215 Topic: animals	Experimental group: Explicit instruction in text structure focused on compare-contrast expository text, emphasis of clue words, generic questions, graphic organisers and close analysis of well-structured text exemplars. Comparison group: The explicit instruction condition was compared with a content lesson programme (which focused on background knowledge, general content discussion, a graphic organiser and animal fact book) and a no instruction group The study was conducted in 4 schools with similar demographics. There were no differences between groups on the performance on a reading comprehension pre-test.	Focus: scientific language Measures included the Woodcock reasoning mastery test (word identification and passage comprehension) and a test to determine whether children can identify a compare/contrast paragraph, generate cue words, questions and comparative statements, and content knowledge of vocabulary concepts. The intervention group scored higher than the content group and the no instruction group on oral and written post-test free summary, prompted summary and pro-con comprehension tasks. There were no differences in content outcome measured between the intervention and content group, with both outperforming the no-instruction group. Effect sizes ranged from 1.36 – 4.40 on outcome measures, considered large effects.
Yeo *et al.* (2020)	Singapore Funded (National Institute of Education) Design: quasi-experiment Age: 9-10 N: 129 (EG=59; CG=70) Topic: temperature and heat	Experimental group: Inquiry-based instruction based on image-to-writing (I2W, the use of images to represent ideas about temperature and heat, and translation of images into text). The I2W unit on temperature and heat was taught over 10 weeks in one all-boys' school and one co-educational school. Control group: the 'image to writing' approach was compared to direct instruction and inquiry (using predict-observe-explain) without multimodal representations in the same schools.	Focus: attainment Measures included an open-ended response test of conceptual understanding and representational competence. The study reports a medium positive effect on conceptual understanding (d=0.42) on pupils in the experimental group, with little difference on the simpler questions and a greater effect on the more complex questions.

Study	Location and sample details	Description of approach	Reported outcomes
Zangori *et al.* (2015)	USA Design: quasi-experiment Age: 8-9 N=116 Topic: the water cycle	Experimental group: the use of instructional scaffolds to teach model-based explanations of the water cycle. The scaffolds consisted of prompts and a diagram on a worksheet. Control group: an unscaffolded condition (fewer instructions; blank box rather than diagram). Non-equivalent control group. Note: connected to the study by Baumfalk *et al.* (2019) in the context of the FOSS water unit.	Focus: working scientifically Measures included modelling tasks completed pre-and post-unit lessons, scored by authors, assessed for epistemic features - components, mapping, explanatory process, scientific principle and sequences. The study found that the embedded scaffolds did not impact four out of five epistemic features in pupils' model-based explanations, but there was a significant ($p=0.02$) effect of scaffolding on pupils' sequencing of events in the water cycle.
Zheng *et al.* (2008)	USA Design: quasi-experiment Age: 9-10 N: 89 Topic: electrical circuits	Experimental group: the experimental group was taught analogical reasoning, a type of model-based reasoning intervention, with and without the use of multimedia. The study took place in a lesson lasting 40-50 minutes. A water system was used as the analogy for electrical circuits. Based on schema theory. Comparison group: Compared multimedia with and without analogy, analogy without multimedia and instruction involving neither multimedia nor analogy.	Focus: attainment, critical thinking Measures included a cognitive test: the Group Embedded Figure Test (Witkin et al. (1971, 2002) and achievement test (recall of conceptual and procedural knowledge, and transfer of knowledge to new situations. The results demonstrated that the multimedia and analogy group outperformed all other groups in recall and transfer, followed by the analogy without multimedia and no multimedia no analogy groups. The multimedia with no analogy group had the lowest mean scores for both recall and transfer. In non-multimedia environment, the group with analogy generally outperformed the group without analogy in both recall and transfer sub-tests. Partial η^2 around 0.03 for recall and transfer effects.

B6 ICT supported and online teaching and learning in primary science

Study	Location and sample details	Description of approach	Reported outcomes
Barak & Dori (2011)	Israel Design: quasi-experiment Age: 9-11 N: 1,335 (EG=926; CG=409) Topics: various	Experimental group: integration of *Brainpop* - animated movies and supplementary activities into lessons at least once a week. Control group: pupils used only textbooks and still pictures for learning science. 15 science teachers from 5 schools participated in 2 hours of CPD. No statistically significant differences between the groups in terms of gender, parents' occupation, and extracurricular activities. About 11% of the pupils declared that their parents' occupation involves a scientific field and 12.8% mentioned that they participated in extracurricular activities related to science..	Focus: critical thinking Measures included a science thinking skills questionnaire. Pupils in the experimental group significantly improved their understanding and implementation of science concepts compared to pupils who used only text-books and still-pictures. This increase was higher for 4th grade pupils than for 5th grade pupils however, no statistical significant difference was found between the two cohorts' net gains . The percentage of correct explanations given by 4th grade ($F(1,623)=7.10$, $p<0.05$) in the intervention group was higher than that of the control pupils. There was no significant difference for 5th grade pupils.

Study	Location and sample details	Description of approach	Reported outcomes
Hodges *et al.* (2020)	USA Funded (National Institutes of Health Science Education Partnership) Design: quasi-experiment Age: 8-11 N: 232 (EG=128; CG=104) Topic: animal body systems	Experimental group: pupils played a Serious Educational Game *Virtual Vet* and participated in guided inquiry activities. Teachers participated in a one day workshop. This was an ICT intervention combined with fixed and growth mindset conditions. Pupils in the control group engaged in hands-on activities such as dissecting an owl pellet and card sorting activities. Teachers were provided with lesson plans so that the same concepts were covered as in the game condition. The pupils in both groups participated in a 45 minute session each week for 9 weeks (approximately 400 minutes in total).	Focus: attainment Tests devised by the research team were used to assess the learning objectives of both conditions and the Lexile framework was used to assess reading ability. The experimental group scored significantly higher in the comprehension test than the control group while spending significantly less time to complete the task. The experimental group's retention score mean was much higher than the score mean of the control group. Pupils in the control group spent more time and cognitive resources while engaged in the activity. In a second study, the serious game only was compared with the serious game combined with guided inquiry and no significant differences were found between groups.

Study	Location and sample details	Description of approach	Reported outcomes
Hu *et al.* (2019)	Taiwan Design: quasi-experiment Age: 10-11 N: 100 (EG1=34; EG2=33; CG=33) Topic: rainfall	Experimental group: taught in a computer room for one class each week for 12 weeks. Pupils given a detailed (EG1) or simple (EG2) stem design as part of a collaborative online questioning system: 'Questioning Supported Thinking and Learning System'. Pupils in the no stem control group (CG) did not receive any prompts.	Focus: attainment, critical thinking Measures: number and quality of questions, quality of science reports and attainment of knowledge. Pupils in EG1 generated the greatest percentage of higher-level questions rated according to Bloom's Taxonomy than the other two groups. EG1 was the only group to generate integration-related questions. Both intervention groups had significantly better performance than the control group in communicating their ideas in science reports. Critical thinking rated using a test: EG1 and the control group had significantly better performance than EG2. EG1 and G3 showed no significant difference. Knowledge tested through a true – false quiz: EG1 and EG2 outperformed the CG but there was no significant difference among the groups.
Hwang *et al.* (2020)	Taiwan Funded (Ministry of Science and Technology, Taiwan) Design:quasi-experiment Age: 10-11 N: 75 (EG1=25; EG2=26; CG=24) Topic: plants	Experimental group: flipped learning, enhanced by problem-posing. EG1 used an online concept map guided problem-posing strategy. EG2 used an online narration-based problem-posing strategy. The control group completed a learning sheet prepared by the teacher. All groups received the same learning materials and completed an activity at home before class. The study took place over 3 weeks.	Focus: attainment Measures of critical thinking and self-efficacy were collected using questionnaires and attainment using a science test. The post-test focused on pupils' comprehension of the course knowledge and to examine whether their critical thinking tendency and self-efficacy were enhanced. Learning achievement test: EG1 outperformed EG2 and the control group on learning achievement but the difference was not significant. Questionnaire for critical thinking tendency (CTT) scale: the results implied that there was no significant difference between EG1 and EG2. However, a significant difference between EG2 and the CG was found. Questionnaire for the evaluation of self-efficacy: no significant differences were found among the three groups on their self-efficacy.

Study	Location and sample details	Description of approach	Reported outcomes
Kim & Olaciregui (2008)	USA Design: RCT Age: 10-11 N: 50 (EG=25; CG=25) Topic: Earth's atmosphere	Experimental group: pupils used a concept map-based information display in an electronic portfolio system. Control group: pupils accessed the electronic portfolio contents in a traditional folder-based 'tree mode' information display. The same teacher monitored both groups and did not provide any information while pupils were engaged in the experiment.	Focus: attainment Measures included information processing performance and memory retention. The experimental group scored significantly higher in the comprehension test than the control group while spending significantly less time to complete the task. The experimental group's retention score mean was much higher than the score mean of the control group . The control group pupils spent more time and cognitive resources while engaged in the activity. The concept map design based on Gestalt principles was identified as particularly beneficial for pupils with low verbal ability,reading skills, or prior knowledge, and for those who speak English as an additional language.
Looi *et al.* (2011)	Singapore Funded (National Research Foundation Singapore) Design: quasi-experiment Age: 8-9 N: 351 Topic: body systems	Experimental group: pupils used mobile technologies in an inquiry context - a mobilised science curriculum. GoKnow MLE (Mobile Learning Environment) was used on a smartphone with apps connected to the internet. The comparison classes were used as control groups and were taught in the traditional way. Note links to inquiry and investigation cluster.	Focus: attainment, attitudes The results show that the intervention class performed significantly better than other classes as measured by traditional assessments in the science subject. 80% of pupils who used the mobile devices thought that the phone helped their learning in and out of class. 62% thought that by using the phone, they understood science concepts better and understood better how they were connected to daily life.

Study	Location and sample details	Description of approach	Reported outcomes
Sun *et al.* (2010)	Taiwan Design: quasi-experiment Age: 10-11 N: 128 (EG=63; CG=63) Topic: the Sun-Moon-Earth system	Experimental group: (3-D) virtual reality model named the Sun and Moon System. 3-D graphics to demonstrate complicated phenomenon and use of a VR model. Pupils in the control group were taught by traditional approaches that used 2-D photographs. Both the treatment and comparison groups observed the real phases of the moon in the sky as homework; however, the treatment group also used the 3-D VR model to assist their observations. Note links to the inquiry cluster.	Focus: attainment, attitudes Researcher-developed tests were used to measure conceptual understanding and a modified questionnaire (Sun, Lin & Yu, 2008) to measure attitude. Results indicate a significant main effect in which the experimental group outperformed the comparison group in understanding the relative positions, motion and phases of the moon. The results indicate that more than two thirds of the pupils agreed or strongly agreed that they would recommend the VR model to peers.
Wang & Tseng (2018)	Taiwan Design: quasi-experiment Age: 8-9 N: 208 Topic: changes of state	Experimental group: pupils in the experimental group undertook virtual interactive laboratory activities involving the use of virtual instruments and objects on a computer before using physical manipulatives. EG1 experienced VMPM - Virtual Material and Physical Materials. EG2 experienced VM alone. One control group (Physical Materials) undertook activities which involved the use of physical instruments and objects. Note links to practical work cluster.	Focus: attainment Science achievement test and two-tier conceptual test used as measures of attainment and conceptual understanding. Scores in a science achievement test to measure pupils' knowledge gains of the science concepts indicated that the VM alone and VM-PM conditions were significantly higher than those of the pupils in the physical materials alone condition.

Study	Location and sample details	Description of approach	Reported outcomes
Ward *et al.* (2013)	USA Funded (Institute of Education Sciences, U.S. Department of Education, the National Science Foundation and the National Institute of Health) Design: quasi-experiment Age: 8-11 N: 1,167 (EG1=83; EG2=69; CG=1015) Topic: various	Experimental group: As a supplement to normal classroom instruction, EG1 was assigned to a conversational multimedia virtual tutor 'My Science Teacher'. EG2 were assigned to an expert human tutor. A total of 4.5 hours of tutoring were given across the 16 sessions. The control group did not receive any supplementary tutoring but did undertake the same modules in class. Class-based instruction for all groups used the Full Option Science System (FOSS) - an inquiry-based science program. Note links to inquiry cluster.	Focus: attainment Standardised tests used to measure learning gain. Results showed that pupils in EG1 and EG2 had significant learning gains relative to pupils in the CG. A moderate effect size was observed for the comparison of EG1 and the CG (reported effect size 0.18) and a larger effect size for EG2 and the CG (reported effect size 0.40). There were no significant differences in learning gains between pupils in EG1 and EG2.
Zacharia *et al.* (2016)	Cyprus Design: quasi-experiment Age 9 N: 48 (EG=24; CG=24) Topic: plants	Experimental group: pupils used mobile devices (90% tablets, 10% smart phones) for data collection about flowers and their functions, namely, taking videos and photos to support their conclusions with data-based evidence. Pupils in the control group used traditional means of data collection including magnifying glasses and sketchbooks. All participants had at least one year of prior experience with the use of mobile devices. Same teacher for both groups. Link to learning outside the classroom.	Focus: attainment Test used to measure conceptual understanding. The mean post-test scores of the experimental group were significantly higher than those of the control group, for the test as a whole and for each part of the test separately. The experimental condition was more conducive to pupils' growth in conceptual understanding than the control condition. Analysis of pupil notebooks revealed statistically significant differences across all checkpoints in favour of the experimental condition.

B7 Language, literacy and text-based studies in primary science

Study	Location and sample details	Description of approach	Reported outcomes
Alexander (2018)	UK Funded (Education Endowment Foundation) Design: RCT Age: 9-10 N: 4,958 (EG=2,492; CG=2,466) Topic: various	Experimental group: Dialogic teaching (maximising pupil talk with the purpose of enhancing pupil engagement and learning) through a CPD intervention. One-year intervention (with one-year pilot and one-year follow-up). Intervention at classroom level. Cost per pupil: £52 per year. Control group: Normal teaching (business as usual). Extensive teacher professional development (118 teachers in experimental group; 90 teachers in control group). All schools met standard poverty criterion of a high proportion of pupil (at least 20 per cent) eligible for free school meals (FSM). English was an additional language (EAL) for about half the pupils involved.	Focus: attainment External evaluation included GL Assessment Progress Tests in English, mathematics and science (GL Assessment is a company that produces tests schools can use to measure progress). Internal evaluation included analysis of questionnaires, interviews and video recordings. Intervention group exhibited higher levels of explanation, analysis, argumentation, challenge and justification. Experimental group pupils on average made two months additional progress compared to CG pupils in English and science, and one month additional progress in mathematics. Effect sizes for science 0.12 for the experimental group, 0.11 for pupils receiving free school meals. For pupils receiving free school meals, the impact was two months across all three subjects.

Study	Location and sample details	Description of approach	Reported outcomes
Bigozzi *et al.* (2011)	Italy Design: quasi-experiment Age: 8-11 N: 172 Topic: combustion, evaporation, and conservation of liquids	Experimental group: pupils participated in individual writing – group discussion – individual writing' (W1-D-W2). They completed their individual writing task before they discussed their observations. Control group: pupils participated in 'group discussion – individual writing' (D-W). They discussed immediately after the observation of the demonstration. EG and CG each half a class of three different grades, comparable SES. The teacher demonstrated the processes to all pupils before splitting class into the control and experimental group. The teacher refrained from making comments during the demonstration, but led the discussion (lasting about 15 minutes) in both groups	Focus: attainment, critical thinking Assessments were made using a coding system for scientific conceptualization and metacognitive thinking applied to pupils' writing. Reported effect sizes 0.34-1.2 for different year groups and their topic activities, on comparing between treatment and control. Comparison of the initial individual writing output (W1, as performed by the treatment group) with group discussion followed by final writing (D-W, as performed by the control group) shows an effect in favour of the latter (reported effect size ranges from 0.02 to 1.10). Personal reflection from an initial individual writing element appears to impact overall learning and metacognition. The study suggests that an initial individual writing element alone is no better than an individual writing element after group discussion - it is the addition of an individual writing element before group discussion and final individual writing that makes the difference.

Study	Location and sample details	Description of approach	Reported outcomes
Bravo & Cervetti, (2014)	USA Design: quasi-experiment Age: 9-11 N: 172 (EG=77; CG=95) Topic: space science	Experimental group: integrated science and language instruction over 40 sessions. The curriculum model incorporates a balance of doing first-hand inquiry, reading in support of inquiry, discussing the investigations and texts, and writing about the investigations and science concepts under study. Control group: Teachers agreed to spend similar time to teach the space science unit according to state standards and using district-adopted curricula, including hands-on science, videos and text as appropriate. No in-person CPD but a detailed teacher guide alongside/embedded in the curriculum materials, with 1-3 EL adaptations in every proposed lesson/session. Two-thirds of the pupils (115 out of 172) are English Language Learners (98% of whom are native Spanish speakers), with 55 in the control group and 60 in the experimental group.	Focus: attainment, scientific language Measures included researcher-constructed tests in science understanding, science vocabulary, and reading comprehension. Experimental group pupils had significantly higher post-test scores for science understanding and vocabulary, but not for reading. ELL pupils benefit more in scientific language development than their non-ELL peers, although there is no significant differential benefit on science reading comprehension. The data are incomplete, so it is not possible to verify the conclusions as drawn by the authors.
Cervetti *et al.* (2012)	USA Design: quasi-experiment Age: 9-10 N=2,019 Cluster allocated: language, literacy and text-based approaches Topic: light and energy	Experimental group: integrated science-literacy unit designed using a curriculum model that engages pupils in reading text, writing notes and reports, conducting first hand investigations, and participating in discussion. Total 40 sessions, 10 sessions for each of four investigations, each lasting 45-60 minutes. Control group: teachers were asked to use the curriculum materials they would regularly use for the state science standards-based topic of light, and match the time and duration of the intervention. Two out of four state topics completely match those of the intervention, and these were used for the research. Experimental and control groups were similar in number (47), size and composition but control group teachers had more experience and education than experimental group teachers.	Focus: attainment, scientific language Measures included identical pre- and post-tests of learning, science understanding, science writing and science vocabulary. Pupils in the experimental group made significantly greater gains on measures of science understanding (reported effect size 0.65), science vocabulary (reported effect size 0.22), and science writing (reported effect size 0.09). Pupils in both groups made comparable gains in science reading comprehension.

Study	Location and sample details	Description of approach	Reported outcomes
Cheng *et al.* (2015)	Taiwan Funded (National Science Council, Taiwan) Design: quasi-experiment Age: 10-11 N: 58 (EG=30; CG=28) Topic: air and combustion	Experimental group: Textbook, with design of text and illustrations modified according to cognitive process principles (minimising extraneous load, enhancing germane load).. Control group: standard textbook published in Taiwan and approved for instruction by the Ministry of Education. Pupils were ranked on prior attainment, after which stratified randomisation assigned them to the experimental or control group. One teacher taught both groups, and their instructional style was to emphasise text when demonstrating concepts, and use of directive/authoritative when guiding experimental work (resulting in pupils following their instructions in place of textbook instructions at times).	Focus: attainment Researcher-developed performance assessment used to measure conceptual understanding. Reported effect sizes are 0.13 for overall performance difference (between EG and CG), conceptual knowledge difference 0.14, retention difference 0.08 and procedural knowledge difference 0.03 (and the latter not significant). Reported effect size for transfer of knowledge to other contexts is 0.09. Enhanced pupils' learning performance in terms of conceptual knowledge, as well as transfer and retention; such improvement did not occur in procedural knowledge. The results are modest, and the study is small-scale. But the principles may be reproducible in other contexts (e.g. other lesson materials containing both text and images) on the basis of the descriptions given, and which are argued on theory.

Study	Location and sample details	Description of approach	Reported outcomes
Connor *et al.* (2017)	USA Funded (Institute of Education Sciences, US Department of Education and the National Institute of Child Health and Human Development) Design: design-based research with RCT Age: 4-9 N: 418 pupils (EG=212; CG=206) Topic: various	Experimental group: Experimental group: For the RCT iteration, CALI – content-area literacy instruction - was implemented, using small group discussions for 15-20 minutes. 3-week unit during literacy block, each unit involving connect, clarify, research and apply lessons. Pupils were taught scientific methods and had the opportunity to conduct experiments and analyse data. Lessons partially scripted. Control group: Literacy core curriculum (Houghton Mifflin) during the literacy block, rated adequate to excellent, with occasional use of expository text but no focused science instruction. Science was taught following the Florida Sunshine State Standards but not during the literacy block. Not observed. Schools from high poverty areas are over-represented in the sample.	Focus: attainment, scientific language Measures included proximal content knowledge assessments for the topic designed by the researchers, standardised vocabulary tests (Woodcock-Johnson III Test of Achievement and Oral Comprehension test). Effect size of CALI on science content knowledge is reported as 2.59, with no change in performance gap between high and low achievers, indicating that CALI can effectively improve pupils' science content knowledge without jeopardising literacy learning. Effectiveness trial has reported overall effect size for science content area knowledge 2.10. Overall reported "Treatment effect size (d) for science is 2.59. During the iterative process of the design of the intervention, developers addressed the potential impact of pupils' incoming language and reading comprehension skills. This was reduced to negligible, i.e. the intervention worked equally well for pupils with all levels of language proficiency. During the effectiveness trial it became clear that pupils with weaker initial language proficiency made greater gains than their peers who had stronger skills at the start.

Study	Location and sample details	Description of approach	Reported outcomes
Goldschmidt & Jung (2011)	USA Funded (Seeds of Science/Roots of Reading Project (commissioned to the National Centre for Research on Evaluation, Standards, and Student Testing) Design: cluster RCT Age: 9-10 N: 2,019 (EG (47 classes) and CG (47 classes) of similar size but CG contained around twice as many EAL pupils Topic: light/energy	Experimental group: Seeds of Science/Roots of Reading project. Pupils in the intervention group had integrated science and literacy integrated instruction which involved pupils in doing, talking, reading, and writing about the characteristics of light, and explicit instruction of comprehension strategies such as making predictions, writing summaries, using nonfiction text structures to find information, and engaging in oral discourse. Intervention delivered over 12 weeks. Control group: content of state science standards related to light, using whatever curriculum materials they regularly use Pre-test confirms the pupils in EG and CG have similar proficiency in writing, despite CG having a higher proportion of EAL pupils.	Focus: attainment, attitudes, scientific language Measures included assessments of science knowledge, science vocabulary, reading comprehension, science writing and student attitudes. Pupils in both conditions demonstrated statistically significant gains; experimental group pupils score higher (reported effect size 0.65); no change in performance gap between high and low achievers due to treatment; treatment results in twice the gain for vocabulary and reading; treatment also has effect on writing outcomes, highly correlated with increased performance on science content knowledge. Teachers report better engagement of pupils in the treatment condition, but the data do not allow analysis of impact of engagement on performance. All subsets of pupils benefit roughly equally. Teacher self-efficacy, but not teacher background and classroom processes, seems to have a differential impact on pupils' science performance. There is some evidence that "teachers not majoring in Early Childhood Education (ECE) tend to have higher student performance" (p29). Inquiry based teachers enhanced treatment effects

Study	Location and sample details	Description of approach	Reported outcomes
Hanley *et al.* (2015) reviewed with Hanley *et al.* (2020)	UK Funded (Education Endowment Foundation) Design: RCT Age: 9-10 N: 1,264 (EG=655; CG=609) Topic: materials, electricity, forces, light and sound, and living things and evolution	Experimental group: Thinking, Doing, Talking Science. CPD programme to increase conceptual challenge and higher order thinking. Two teachers per school attended five training days over the school year. CPD covers philosophy, techniques and activities (not in the specific use of specific resources), and teachers are expected to complete 'gap tasks'. Teacher interaction is encouraged during and after CPD, and therefore a whole school is either in EG or in CG as multiple teachers may be involved. Control group: business as usual with wait-list teachers who were offered the training intervention at a later date, and their pupils. Note links to inquiry and collaborative learning clusters.	Focus: attainment, attitudes Measures include a pencil-and-paper science test covering a range of topics and question types and an attitude questionnaire. Statistically significant effect on attainment in favour of experimental group, with an effect size of 0.22. Overall, Year 5 pupils in schools using the approach made approximately three additional months' progress" (p4). Some indications that the approach had a particularly positive effect on pupils eligible for free school meals (reported effect size 0.38, but not statistically robust), on girls (reported effect size 0.32) and on pupils with low prior attainment (reported effect size 0.33).
Henrichs & Lesemann (2014)	The Netherlands Funded (Dutch National Platform Science and Technology) Design: RCT Age: 5-6 N: 241 Topics: air pressure and reflection	Experimental group: Teachers received training and materials to conduct small group discussion with a strong focus on academic/scientific language, integrating science and language learning. CPD on academic language lasting three hours, with materials and link to website provided (no longer available). Control group: waitlist teachers were offered the training after completion of the study; they used the same tasks and topics (from a Dutch national research programme investigating children's early science talents and curiosity) as the experimental group teachers but without any training	Focus: scientific language Observations measured incidences of on-task utterances and words, scientific reasoning and lexical diversity. Where teachers are trained in the appropriate use of academic language for science education, pupils' conceptual understanding and scientific language benefit. Reported effect sizes for lexical diversity, general academic words and domain-specific words: 0.11-0.80, depending on scientific topic. The authors report an overall effect size 0.63 for the pupils, and an overall effect size 0.96 for teachers.

Study	Location and sample details	Description of approach	Reported outcomes
Jay *et al.* (2017)	UK Funded (Education Endowment Foundation) Design: RCT Age: 9-10 N: 1,233 (EG=614; CG=619) Topic: various	Experimental group: dialogic teaching approach. Intervention emphasising dialogue through which pupils learn to reason, discuss, argue and explain. Intervention delivered over a year. Cost reported (£52/pupil/year). Professional development for teachers 3 days/teacher, 4 days/mentor, 1 day/headteacher. 6 x planning and review meetings in school. Schools in the control group were asked to engage in 'business as usual' although it was not stated what this was. Note links to Alexander (2018).	Focus: attainment Science attainment measured using the Progress Test in Science. The post-test marks for the experimental group in science were higher than the control group, showing a modest effect size (reported effect size 0.12). For pupils eligible for free school meals, there was a modest effect size for science (reported effect size 0.11). This was not significant. Outcomes from post-intervention case studies indicated that teachers reported positive effects on pupil engagement and confidence. They also reported an increase in the quantity and quality of talk.
Kim *et al.* (2021)	USA Funded (Chan Zuckerberg Initiative) Design: cluster RCT Age: 6-8 N: 5,494 (EG=2,886; CG=2,608) Topic: How do animals survive in their habitat?	Experimental group: Science as well as social studies lessons through Model of Reading Engagement. Teachers integrated thematic lessons, concept mapping, and interactive read-alouds of conceptually related informational texts followed by collaborative research and writing tasks. Initial 3-hour CPD for teachers, followed by daily video clips to explain lesson resources. Monthly meetings for literacy facilitators. Delivered over 5-10 weeks (20 lessons). Control group not described in any detail, apart from instruction minutes devoted to certain aspects of teaching, in science, reading, and social studies. Experimental and control groups comparable on measures of SES, EAL and ethnicity	Focus: critical thinking, scientific language Semantic association task (Read, 1998, 2004) used to measure vocabulary depth, Measure of Academic Progress (MAP) Primary Grade Reading (Northwest Evaluation Association, 2011) used to measure comprehension, argumentative writing assessed using a rubric. Positive effects of intervention on pupils' science vocabulary knowledge depth (reported effect size 0.50) and argumentative writing (reported effect size 0.24), but not on domain-general measures of reading comprehension (reported effect size 0.11). First-grade classes not receiving the intervention in schools where the second-grade class **is** receiving the intervention, function as control group for the reverse situation. Overspill of intervention ideas/opportunities from intervention classes to non-intervention classes in the same school, is not discussed. Investigation of vocabulary knowledge depth as a mediating factor: vocabulary knowledge is a mediating factor in reading and argumentative writing outcomes.

Study	Location and sample details	Description of approach	Reported outcomes
Kitmitto *et al.* (2018)	UK Funded (Education Endowment Foundation) Design: two-arm cluster RCT Age: 9-10 N: 8,996 (EG=3,961; CG=4,054 Topics: materials, electricity, forces, light and sound, and living things and evolution	Experimental group: Thinking, Doing, Talking Science CPD intervention training teachers to develop and teach challenging, inquiry-based science lessons that incorporate more practical activities, deeper thinking and discussion, and sharply focused recording. Duration of CPD was 4 days (shorter than in intervention evaluated by Hanley et al., 2015) and a train-the-trainer model was used. Teachers provided with support materials. Gap tasks including trying activities: e.g. an odd one out and practical prompt. Control group: waitlist group of teachers, and their pupils. Teachers in the control group were asked to record their CPD experience as a measure of business as usual. Cost £29/pupil/year.	Focus: attainment, attitudes Measures include a science knowledge questionnaire, interest in science index and pupil survey. Study found no evidence that TDTS had an impact on pupils' science knowledge attainment, on average. Reported effect size 0.01. Among pupils receiving free school meals, those in TDTS schools made a small amount of additional progress compared to those in other schools but not at a statistically significant level. Reported effect size 0.05. TDTS led to small increases in pupil interest in science and self-efficacy for science. Reported effect size for interest 0.12, and for self-efficacy 0.09.
Lai & Chan (2020)	Taiwan Design: quasi-experiment Age: 10-11 N: 118 (EG=59; CG=59) Topics: the influence of heat on matter and air and burning	Experimental group: Reading three science trade books, experiments, mind maps, discussions. Intervention delivered over 9 weeks (27 sessions) Control group: traditional instruction, with pupils given optional access to the same trade books as those in the EG, but these books were not referred to explicitly in lessons There is no further mention of the employment of the trade books in the control group, even though those books were accessible to the pupils.	Focus: attainment, attitudes Science achievement test and attitudes towards science scale used to measure impact of the intervention. Science Achievement Test scores higher for experimental group than control group ($F=4.921$, $p<0.05$). Integrating science trade book reading improved pupils' science understanding.

Study	Location and sample details	Description of approach	Reported outcomes
Llosa *et al.* (2016)	USA Funded (National Science Foundation) Design: RCT Age: 10-11 N: 6,673 (EG=2894; CG=3345) Topics: nature of science, earth and space science, life science, and physical science	Experimental group: One year Promoting Science among English Language Learners (P-SELL) curriculum. Standalone, year-long science curriculum designed to promote students' scientific inquiry and understanding while providing language development strategies. Organised around big ideas. Teachers provided with CPD (5 days), educative materials and resources. Control group: district-adopted science textbooks, i.e. Interactive Science by Pearson (district A), National Geographic Science (district B), Science Fusion by Houghton Mifflin Harcourt (district C). All time and duration comparable to that for EG. Experimental and control group numbers combine to a number for analysis which is smaller than the number of pupils overall; 23.1% of pupils were or had been classified as English language learners. Note link to Maerten-Rivera et al. (2016).	Focus: attainment State science assessment used as an outcome measure and a researcher-developed science assessment allowed for a pre-measure of science achievement. Positive impact on pupils' science achievement in both state assessment (reported effect size 0.15) and researcher developed science assessment measures (reported effect size 0.25). Effect sizes are larger for ELLs than non-ELLs (0.35, 0.41, 0.28, 0.24 resp. for the four classifications, starting at ELLs) on the researcher-developed measures. Differentially positive impact on ELL pupils on the researcher-designed test compared to equivalent pupils in the control group. On the state assessment, non- or former English language learners performed significantly better in the treatment group compared to control group, with positive but non significant differences for ELLs or recently classified ELLs. Pupils' initial language proficiency does not seem to be a moderating factor.

Study	Location and sample details	Description of approach	Reported outcomes
Maerten-Rivera *et al.* (2016)	USA Funded (Institute of Education Sciences, Department of Education) Design: cluster RCT Age: 10-11 N: 20,879 (EG=9,957; CG=10,722) Topics: nature of science, earth and space science, life science, and physical science	Experimental group: One-year dedicated Promoting Science among English Language Learners (P-SELL) curriculum, standards-based, inquiry-oriented (moving progressively from teacher- to pupil-directed inquiry), and ELL-focused (explicit guidance in science concepts, inquiry and language). Integrates big ideas with bodies of knowledge. Teachers' guide and resources provided. Extensive CPD (5 days in years 1 and 2 and 1 day in year 3). Control group: district-adopted science curriculum, which includes kits of science supplies for hands-on inquiry at three different levels of complication: "directed", "guided" and "full" inquiry, as well as explicit links to other subjects (including literacy, although the activities suggested were not designed to specifically support ELLs). Inquiry and literacy activities are not always clearly linked to the science concepts under development.	Focus: attainment Science proficiency measured using state science assessment. State reading assessment also used. Across all 3 years, the percentage of proficient students was greater in the treatment group than the control group, with female pupils, Black pupils, pupils who are English language learners and those eligible for free or reduced lunch having the lowest levels of proficiency. Impact of the programme increases with time after first implementation: no difference between experimental and control in first year, small but not yet significant difference in second year, significant difference in third year. Teachers need multiple years to adjust their teaching practices to become effective before impact can be seen in pupils' proficiency.
Vitale & Romance (2012)	USA Funded (National Science Foundation) Design: quasi-experiment Age: 6-8 N: 363 (EG=180; CG=183) Topics: life, Earth, and physical sciences	Experimental group: Science IDEAS, 45 minutes each day as part of literacy instruction, complementing the district reading/language Arts programme. Instruction included groups in observing, exploring and discussing. Teachers received CPD (4.5 days total). Control group: the regular district science programme, in addition to the district reading/language Arts programme. Experimental and control group demographically similar. Grade 3-5 teachers in both control and experimental group schools were also involved with *Science IDEAS* and operated as mentors for their grade 1-2 colleagues.	Focus: attainment, scientific language ITBS Reading Comprehension and Science subtests (level 7 for grade 1, level 8 for grade 2) used for measures of language and science attainment. Treatment/intervention were statistically significant in favour of the grade Science IDEAS classrooms for both ITBS Science and Reading, with ethnicity as a moderating factor for science attainment.

Study	Location and sample details	Description of approach	Reported outcomes
Zwiep & Straits (2013)	USA Design: quasi-experiment Age: 4-11 N: 3,347 (EG=2122; CG=1225) Cluster allocated: language, literacy and text-based approaches	Experimental group: Blended inquiry science (5E format) with English Language Development, including careful consideration of vocabulary, some front-loaded, some embedded in the inquiry activities. Teachers received 2-week intensive CPD and participated in site-specific learning collectives. Control group: the district's established English Language Development programme, alongside "very little, if any, instruction in science" (p1317) Experimental and control groups comparable on pupil demographics and previous attainment.	Focus: attainment Measures of language achievement gathered using the California English Language Development Test (CELDT) and the California Standards Test (CST) in English Language Arts (ELA). Science attainment measured using the Science CST and a science content assessment developed by WestEd. Small effects on both science and language tests, analysed in many ways with sub-skills also analysed and found benefiting; no harm to English language development to those in the programme. Statistical analyses are based on varying sample sizes, depending on the length of time a pupil was exposed to the intervention, and the availability of standard state test results.

B8 Learning outside the classroom in primary science

Study	Location and sample details	Description of approach	Reported outcomes
Glick & Samarapungavan (2008)	USA Design: quasi-experiment Age 9-10 N:30 (EG=19; CG=11) Topic: wolf behaviour	Experimental group: learning outside the classroom. School trip with use of field notebook and activities before and after the school trip. Linking learning in different contexts. Control group: School trip without supporting activities. In both cases, the field trip was led by experienced staff at Wolf Park providing a structured field trip.	Focus: attainment Knowledge test used pre-and post-visit. Univariate tests on knowledge about wolves and ecology more generally revealed statistically significant differences between the experimental and control groups in favour of the experimental group on both the post-test ($F=8.177$, $p<0.01$, $M=17.63$ and $M=12.50$) and interview ($F=9.851$, $p<0.005$, $M=31.95$ and $M=25.13$). Note very small sample.
Mills & Katzman (2015)	USA Funded (National Science Foundation) Design: quasi-experiment Age: 10-11 N: 151 Topic: Who wants to be a scientist?	Experimental group: learning outside the classroom. Field trip to science research and learning centre (including an introductory movie, toured control room of science facility, experienced hands-on activities, and interactive displays) and opportunity to interview scientists (30 min) based on a list of suggested questions. Control group: Only field trip to science research and learning centre (as above) but without interviewing scientists.	Focus: attitude (desire to become a scientist, participation in science activities) Measures of Possible Science Selves, Creative Tendencies and Career Interest gathered using questionnaires. For the experimental group, from pre- to post-test a significant increase in the *desire* to become a scientist and *participation i*n science activities (both $p<0.001$) occurred with large (Cohen's $d=0.67$) and medium effect size ($d=0.55$) respectively. For the control group a significant increase in the *desire* to become a scientist and *participation i*n science activities (both $p<0.05$) occurred with a small Cohen's d (0.29 and 0.28 respectively). There was a significant difference in the *desire* to become a scientist ($p<0.001$) with a higher desire for the experimental group (effect size Cohen's $d=0.40$, so small-to-medium). No significant difference for the participation in science activities between the experimental and control groups.

Study	Location and sample details	Description of approach	Reported outcomes
Piila *et al.* (2021)	Finland Funded (European Union) Design: quasi-experiment Age: 10-12 N: 364 (EG=274; CG=90) Topic: Mars colonisation	Experimental group: Learning outside the classroom. STEAM-based inquiry strategy, but Arts component in STEAM not detailed. Use of 'scientific technology' (e.g. IT applications, augmented reality and international exchange platforms) with considerable out-of-school components (visits to planetarium, museum and science centre, environmental education workshops) for designing MARS colonisation unit, drawing on everyday creativity to develop thinking skills. Group tasks with peer discussion intended to develop 'assisted metacognition'. Activities follow national core curriculum unit 'Transversal competencies'. Control group: Business as usual within stated curriculum unit with variable (unspecified) teaching approaches. Link to cross-curricular approaches.	Focus: attainment For thinking skills, general science knowledge was tested (not testing understanding of Mars or space) measuring if experimental group pupils had an advantage in answering items that required transfer of knowledge, science thinking or reasoning. The experimental group improved their scores for science thinking skills significantly ($p<0.001$, effect size 0.15) whilst control group pupils improved significantly but less so ($p=0.04$, effect size 0.05). The intervention results in significantly increased learning outcomes for medium and high achieving groups, but not for the low achievers. The intervention results in similar significantly increased learning outcomes for girls and boys (CG: $p=0.02$, effect size 0.01; EG: $p=0.02$, effect size 0.02). EG girls and boys improved significantly ($p<0.0001$, effect size 0.15 and $p<0.001$, effect size 0.14, respectively), but only the CG boys improved significantly ($p=0.004$, effect size 0.21). Significantly increased learning outcomes for high achieving groups ($\chi^2=4.70$, $p=0.03$),
Scott & Boyd (2016)	UK Funded (Esmée Fairbairn Foundation) Design: quasi-experiment Age: 9-11 N: 379 Topic: ecology	Experimental group: learning outside the classroom. Classroom teaching AND ecological fieldwork. Habitats include school playing fields and gardens; a school pond; local woodland, the hedgerow along a local bridleway, and the local rocky shore. Task to identify as many organisms as possible, photograph species of interest, make notes about appearance and location, write down questions. Control group: same ecological content taught in classroom setting. No fieldwork.	Focus: attainment Literacy and science pre-tests and post-fieldwork assessment tasks used to measure knowledge. Whereas pupils in the experimental and the control group made significant improvement in their science (ecological) knowledge, pupils in the intervention classes exhibited significantly higher levels of post-test achievement than those in the comparison classes ($F=24.01$, $p<0.001$).No differences by gender.

Study	Location and sample details	Description of approach	Reported outcomes
Wells *et al.* (2015)	USA Funded (US Department of Agriculture) Design: quasi-experimental design Age: 5-12 N=3,061 (EG=1622; CG=1439) Topic: gardening	Experimental group: learning outside the classroom. Healthy Gardens, Healthy Youth study. Raised-bed or container garden kits; services of cooperative extension educators (outsiders); educational toolkit for educators; learning resources for school teachers; garden implementation guide. Control group: business as usual but no information on strategy being used or topics covered.	Focus: attainment Science knowledge was measured using a 7-item questionnaire focused on nutritional and plant science. Intervention, consisting of both classroom- and garden-based lessons, had a positive effect on children's knowledge of plant science and nutritional science. The gain in the experimental group was significantly ($p<0.001$) greater than the gain in the control group over the two-year study. Higher implementation fidelity corresponded to significantly ($p<0.001$) higher treatment effects.
Wünschmann *et al.* (2017)	Germany Design: quasi-experiment Age: 8-10 N: 65 (EG=23; EG2=18; CG=24) Topic: reptiles and amphibians	Experimental Group: learning outside the classroom. EG1 (Reptilium group): one lesson at school, a visit to reptile and amphibian zoo with a school-based debrief. EG2 (School group): class-based lessons covering the same reptilium content. Control group: regular science lessons unrelated to reptiles but on 'wondering about biological diversity'. Small sample sizes makes extensive statistical treatment a little artificial.	Focus: attainment, attitudes German adaptation (Wilde et al. 2009) of the intrinsic motivation inventory (Deci & Ryan 2003) used to measure motivation. School grades used to measure knowledge. The two treatment groups (school group and Reptilium group) differed significantly in the science knowledge about reptiles post-test (IRT factor scores, t test, f1: $t=-4.05$, $p<0.001$; f2: $t=-4.40$, $p<0.001$). The differences in the post-test (item response factor scores) were highly significant (t test, f1: $t=-6.19$, $p<0.001$; f2: $t=-4.54$, $p<0.001$), which suggests that the treatments led to substantially greater knowledge. The results were similar for the post-test and follow-up test.

B9 Practical work in primary science

Study	Location and sample details	Description of approach	Reported outcomes
Cvjetićanin *et al.* (2015)	Serbia Funded (Ministry of Education and Science of Serbia) Design: quasi-experiment Age: 9-10 N: 136 (EG=68; CG=68) Topic: physical and chemical properties of materials	Experimental group: practical work with hands on experience with structured experiments on physical and chemical properties of materials, leading to clear results. Control group: the teacher demonstrated the same experiments.	Focus: attainment Test consisting of 12 tasks which evaluated the six levels of knowledge: remembering (retrieving, recalling, or recognizing knowledge from memory). The experimental and control groups were equally successful in improving their knowledge about materials at the cognitive level of remembering (t=1.762, p=0.08), understanding (t=0.742, p=0.43) and application (t=0.868, p=0.38). The difference in scores between groups was not significant. The experimental group (hands-on experience) achieved better results than the control group (demonstration) on level of analysis (t=2.329, p=0.021), evaluation (t=6.764, p<0.001) and creation (t=10.157, p<0.001). These significant differences are maintained in the delayed post-test.
Dankenbring & Capobianco (2016)	USA Funded (National Science Foundation) Design: quasi-experiment Age: 10-11 N: 67 (EG=37 drawn from two half classes; CG=30 as above) 2 classes, Topic: Sun-Earth Relationships	Experimental group: science Learning through Engineering Design (SLED). Teacher-led classroom activities starting with discussions of everyday questions about day/night and the seasons, complemented with design task (design structure providing shade in summer and winter) based on a design brief. Learning materials provided. CPD lasting two weeks. Control group: continued teacher-directed science activities	Focus: attainment Draw-and-explain item used to measure pupils' understanding of seasons. No significant difference between the learning gains of the experimental and control groups.

Study	Location and sample details	Description of approach	Reported outcomes
Durmus & Bayraktar (2010)	Turkey Design: quasi-experiment Age: 9-11 N:104 (EG1=34; EG2=35; CG=35) Topic: matter and change	Experimental group: practical work. EG1: provision of eight self-developed conceptual change texts on 'Matter and Change' focused on pre-identified misconceptions for the class. EG2: provision of eight hands-on structured experiments using Predict-Observe-Explain approach leading to teacher guidance for conclusions. Teachers were trained to use the texts (EG1) and the lab activities (EG2) Eight sessions Control group: 'traditional lecture method'	Focus: attainment Conceptual understanding measured using a Matter Concept Test consisting of eight open ended questions. The results of the study (descriptive statistics, in percentages) showed that both experimental groups were more successful than the control group in overcoming misconceptions permanently (no evidence that difference is significant). There is no obvious difference between the results from EG1 and EG2. Study suggests that conceptual change texts are as effective as hands-on laboratory experiments.
Leuchter *et al.* (2014)	Switzerland Design: quasi-experiment Age: 4-9 N: 288 (EG=244; CG=44) 17 classes, 15 EG, 2 CG Topic: floating and sinking	Experimental group: practical work. Structured and problem-based learning environment, allowing everyday hands-on experiences. Intent on conceptual change: common misconceptions are elicited; conceptual restructuring focusing on the construct of 'material kind' as predictor of floating and sinking; developing to basic notion of buoyancy. Control group: business as usual, with no teaching on floating and sinking.	Focus: attainment, critical thinking Pre-, post- and delayed post-tests of conceptual understanding The experimental group showed a significantly larger increase in conceptual understanding of floating and sinking of solid bodies in the post-test ($F=46.98$, $p<0.01$, $\eta^2=0.17$) and the delayed post-test ($F=24.05$, $p<0.01$, $\eta^2=0.13$), Similarly, the experimental group showed significantly more gains than the control group in their reasons for their classifications of object behaviour. For the experimental group, there was no interaction effect between learning gain and school level ($p=0.91$). Kindergarten children and lower primary children showed comparable learning gains.

Study	Location and sample details	Description of approach	Reported outcomes
Meyer *et al.* (2016)	Germany Design: quasi-experiment Age: 10-11 N: 166 (EG1=57; EG2=53; CG=46) Topic: animals	Experimental group: practical work. EG1: Short term contact and structured hands-on experiments (worksheets) with harvest mice in the classroom. EG2: Long term contact and structured hands-on experiments (worksheets) with harvest mice in the classroom. Tasked with care of harvest mice in the class. Control group: no mice in the classroom. Watched same behavioural experiments with harvest mice in short films on laptops. All groups received the same four teaching units, and worked in groups of 5-6 pupils. 6 classes (2/group).	Focus: attitudes Test instruments were adapted versions of the Flow Short Scale (FSS, Rheinberg et al., 2003) and the Intrinsic Motivation Inventory (IMI, Ryan, 1982). The control group produced significantly lower scores for both engagement/flow ($F=18.16$, $p<0.001$, $\eta^2=0.27$) and motivation ($F=3.13$; $p<0.01$; $\eta^2=0.07$).A significant difference was found between students with short-term versus long-term contact for the motivation subscales (interest/enjoyment, perceived competence, and perceived choice): only pupils in the experimental group with long term experience of mouse keeping showed significant differences for each of the subscales.
Ünal & Aral (2014)	Turkey Design: quasi-experiment Age: 5-6 N: 42 (EG=22; CG=20) Topic: not stated	Experimental group: Practical work. 'Experiment-Based Education Program': 20 structured everyday experiments requiring basic scientific process skills introduced by science-related activities (music, play, stories) e.g. problem solving – the best way to remove glitter powder from their hands. Control group: business as usual, but no details.	Focus: working scientifically Problem Solving Scale in Science Education developed by researchers used to measure problem-solving skills. There was a significant difference between the scores of pupils in the experimental and control groups (t:11.9, $p<0.05$) post-test suggesting the intervention was effective in supporting the pupils' problem-solving skills. The problem-solving skills gained through the Experiment Based Education Program had persistence over time ($p<0.05$).

B9 Inquiry and investigation in primary science

Study	Location and sample details	Description of approach	Reported outcomes
Chen & She (2015)	Taiwan Funded (National Science Council, Taiwan) Design: quasi-experiment Age: 9-10 N: 116 (EG=62; CG=54) Topics: sound, heat transfer, and stars	Experimental group: scientific inquiry programme with framework (=driving questions) for pupils to formulate hypotheses, identify variables, make conclusions based on experimental data, generate evidence-based scientific explanations. The intervention was delivered in one school over 3 months. The integration of scientific reasoning into inquiry learning is manifested by the requirement to answer and justify responses to these driving questions in writing. Control group: Same scientific Inquiry programme and the same set of driving questions but they do not need to be answered and justified in writing, so not an *explicit* scientific reasoning component.	Focus: attainment, critical thinking working scientifically Measures included a scientific concepts test (SCT), scientific concept-dependent reasoning test (SCDRT), and scientific concept-dependent inquiry test (SCDIT) developed by researchers. The experimental group outperformed the control group, in the post and delayed post-test of scientific concepts (($F=4.56$, $p=0.04$, effect size $\eta^2=0.04$ (S); and $F=16.21$, $p<0.001$, effect size $\eta^2=0.12$ (M), respectively). The experimental group had the same significant advantage for scientific concept-dependent reasoning ($F=19.09$, $p<0.001$, effect size $\eta^2=0.14$ (L) for post to retention scores), and for scientific inquiry ($F=6.15$, $p=0.02$, effect size $\eta^2=0.05$ (S); and $F=23.26$, $p<0.001$, effect size $\eta^2=0.17$ (L), for pre-post and post-retention change respectively). The experimental group generated a significantly greater number of testable hypotheses ($F=9.88$, $p=0.002$, effect size $\eta^2=0.08$ (M)), and correct evidence-based scientific explanations ($F=15.95$, $p<0.001$, effect size $\eta^2=0.14$ (L)) than did the control group.

Study	Location and sample details	Description of approach	Reported outcomes
Chen *et al.* (2019)	Taiwan Funded (Ministry of Science and Technology, Taiwan) Design: quasi-experiment Age: 9-10 N: 68 (EG=32; CG=36) Topic: various	Experimental group: Modified Argument-Driven Inquiry (MADI) in laboratory (see above for steps). Standard semi-guided Inquiry (everyday problem presentation; pupils formulate researchable questions; teacher helps to select suitable questions; groups do experiments; groups present findings/claims; teacher encourages argumentation-based discussion). This intervention provided teacher professional development and learning materials for solving several challenges based on play activities. Control group: regular science lessons including teacher-guided hands-on experiments in class without engaging in experimental design and hypothesis generation.	Focus: attitudes, critical thinking Measures: Elementary School Student Questionnaire used to measure engagement in science learning and argumentation abilities. Significant positive effect on ELS for the experimental group compared to the control group for the total engagement in learning science (η^2=0.10), especially emotional engagement (low anxiety (η^2=0.18), enjoyment (η^2=0.14), pleasure (η^2=0.13) but no significant effect on cognitive engagement (motivation). Significantly positive effect on argumentation skills (η^2=0.12) and for each argument dimension: claim (η^2=0.14); evidence (η^2=0.11); warrant (η^2=1.00); rebuttal (η^2=0.09). Both effects are particularly prominent in 2nd semester so improvement in both learning outcomes needs time. No significant changes in engagement in science for boys and girls in both groups. Significant effect on argumentation skills for both experimental group girls (η^2=0.44) and boys (η^2=0.39) and for control group boys (η^2=0.66).
Di Mauro & Furman (2016)	Argentina Design: quasi-experiment Age: 9-10 N: 60 (EG=30; CG=30) Topic: experimental design	Experimental group: inquiry. Bespoke guided-inquiry curriculum unit focusing on experiments based on cross-domain (related to different science disciplines) problems with low (science) conceptual load (=everyday knowledge). Each simple experiment, promoting the ability of experimental design, allows for clear identification of the dependent and independent variables. Control group: business as usual, based on teacher-led text interpretation and occasional teacher demonstrations.	Focus: working scientifically. Researcher-designed measures of working scientifically: tests consisting of three problems set in everyday contexts requiring open answers. Experimental design skills significantly improved for the experimental group (χ^2=27.98, $p<0.001$) but remained constant for the control group. None of the groups showed any difference between post- and delayed post-test ($p=0.47$) so improved experimental design skills were maintained.

Study	Location and sample details	Description of approach	Reported outcomes
Lai (2016)	Taiwan Design: quasi-experiment Age: 8 N: 106 (EG=53; CG=53) Topic: air	Experimental group: Inquiry approach. iPod (invitation-Prediction-Operation-Discussion) inquiry teaching method used along with multimedia (e.g. science toys, video animations) justified as conceptual change strategy. Little detail on intervention – only principles. Control group: use of standard (named) textbook in Taiwan	Focus: attainment, attitudes The research instruments consisted of an Air Concepts Comprehension Test and the Scientific Attitude Scale. Achievement on air concepts was significantly higher for the experimental group compared to the control group ($F=48.50$, $p<0.001$). Scientific attitude was significantly higher for the experimental group compared to the control group ($F=5.66$, $p<0.05$).
Lai *et al.* (2018)	Taiwan Funded (Ministry of Science & Technology) Design: quasi-experiment Age: 10 N=56 (EG=29; CG=27) Topic: insects	Experimental group: inquiry. Computer-assisted Self-Regulated Science Inquiry Approach (SRSIA). This is 5E science inquiry cycle + self-regulation (S-R) steps. So self-regulation is being practised. Control group: conventional science inquiry approach using 5E model. Both groups do the same Science Inquiry activities	Focus: attitudes, working scientifically Questionnaire of information seeking strategies, self-efficacy, and self-regulation used to measure metacognition and teacher designed test used to measure knowledge. Comparing the experimental and control group data, the intervention results in significantly higher (i) learning achievement ($\eta^2=0.23$) ; (ii) information seeking tendency ($\eta^2=0.10$) independent of self-regulation levels; (iii) formal query tendency ($\eta^2=0.14$) but not their informal query tendency; and (iv) self-regulation ($\eta^2=0.87$) in particular for initial low-self regulators, and not for initial high self-regulators. Experimental group pupils with initial higher self-regulation improve their learning achievement more than those with low self-regulation ($\eta^2=0.19$) with no differentiation in the control group.
Li *et al.* (2016)	China Funded (National Natural Science foundation of China) Design: quasi-experiment Age: 10	Experimental group: Inquiry. Using commonly used science pedagogy with LEGO bricks for problem solving, with teaching steps: "creating situations, analysing problems, building a prototype, and testing a prototype, with an engineering design-based pedagogy. Thus, the second step was divided into three more detailed sections: (a) describe the problems on a design card to analyse the constraints, (b) draw a design diagram on a	Focus: critical thinking Two physics tests used to measure attainment and Problem-Solving Ability Self-Checking Questionnaire (Li, 2003) used to measure problem solving. Both groups show large gains in science attainment, but the results show no difference between the groups (effect size $r=0.013$). Therefore, the intervention has no impact on science attainment.

Study	Location and sample details	Description of approach	Reported outcomes
	N:30 (EG=15; CG=15) Topic: engineering design	design card, and (c) decide the optimal solution after comparing potential solutions with a score sheet. Control group: Only using commonly used science pedagogy with LEGO bricks for problem solving with same four steps.	The experimental group shows significantly more gains in problem-solving ability ($p<0.05$, effect size $r=0.40$). Note small samples make refined statistics a little artificial.
Lin *et al.* (2009)	Taiwan Funded (National Science Council of Taiwan) Design: quasi-experiment Age: 10-11 N: 92 (EG=31; CG=61) Topic: various	Experimental group: Inquiry-based learning (hypothesis making; designing experiments; draw conclusions from data; solve problems) added, using play activities rather than everyday contexts. General emphasis on questioning and responding during the inquiry process. Activities and materials provided. Control group: business as usual, with structured hands-on activities from textbook without hypothesising or experimental design.	Focus: attitudes, critical thinking, working scientifically Attitudes measured using learning environment questionnaire with seven subscales (student cohesiveness, teacher support, involvement, investigation, task orientation, cooperation, and equity). No significant difference found for the experimental group on 5 of the 7 scales. A significantly lower perceived teacher support than (only) one of the two control classes ($p<0.05$) and a significantly higher perceived involvement in their learning than the other control class ($p<0.05$). Within the experimental group, the high-quality level question asking students significantly outperformed the low-quality level question asking pupils in the quality of designing experimental procedures ($p<0.05$, $t=2.79$), but not in making hypotheses.
Letina (2016)	Croatia Design: quasi-experiment Age: 9-10 N: 333 (EG=169; CG=164) Topic: life conditions See Letina (2020) - same sample	Experimental group: Inquiry-based teaching strategy and the use of scientific methods and procedures: observation, description, comparison, data collecting, logging and displaying data, data reasoning and interpretation, hypotheses definition, research planning, experimentation, individual reading of literature and designing of research reports. Prepared according to specially prepared plans based on the inquiry teaching methods.16 classes, 8 CG, 8 EG CG: Traditional, lecture-based teaching of special lesson plans developed for the same content of the Curriculum for Science and Social Studies.	Focus: attainment In fact, levels of scientific literacy have been tested. Other - the dependent variable in this research is students' scientific competence (their science knowledge and skills for solving scientific problems of different complexity and their capability for argumentative reasoning of the obtained solutions). Outcomes: results show that the EG achieved significantly higher total scores for the sum of levels of scientific literacy ($F=164.92$, $df=1$, $p<0.01$) as compared to CG.

Study	Location and sample details	Description of approach	Reported outcomes
Letina (2020)	Croatia Design: quasi-experiment Age: 9-10 N: 333 (EG=169; CG=164) Topic: life conditions	Experimental group: inquiry-based teaching of primary science (see Letina (2016)). Control group: traditional, lecture-based teaching of the same curriculum content. See Letina (2016) - same sample.	Focus: attitudes Learning to learn competence was measured using an adapted questionnaire developed by the Institute for Social Research in Zagreb, Center for the Research and Development of Education. The learning to learn competence of students from the experimental group was better to a statistically significant degree compared with students from the control group in the same situation ($F=6.31$; $df=1$; $p<0.05$).
Polikoff *et al.* (2018)	USA Funded (Mattel Children's Foundation) Design: cluster RCT Age: 9-10 N=1,651 (EG=894; CG=818) Topic: speedometry	Experimental group: Practical work, discussion and cooperation. Using 5E strategy (quite guided and linear) aligned with NGSS and CCSS in Science, Maths and English Language. Specific Speedometry curriculum materials, including videos, activities, worksheets, sets of Hot Wheels cars and tracks. Brief CPD (set of videos) for teachers to become familiar with Speedometry curriculum. Control group: business as usual, NGSS and CCSS with only one in three classes addressing the Speedometry curriculum. 53 classes, (high % ELL, high % FRL)	Focus: attainment, attitudes Student Content Knowledge Assessment used to measure conceptual knowledge and student interest and emotions survey used to measure affective responses to the curriculum. The intervention led to a 0.48 standard deviation increase in student content knowledge ($p<0.001$) versus the control. The Speedometry curriculum led to a 0.42 standard deviation increase in student interest but not statistically significant at $p<0.05$ level. Speedometry led to a 0.30 standard deviation significant increase in positive emotions such as excitement ($p<0.05$) and decrease in negative emotions such as boredom with 0.30 sd decrease ($p<0.05$); frustration with 0.26 sd decrease ($p<0.05$); and confusion with 0.25 sd decrease ($p<0.01$). The curriculum did not have a significant impact on surprise or curiosity. The intervention did not have a differential impact on pupils' content knowledge, their interest nor emotions based on gender, ELL and SPED status.
Schalk *et al.* (2019)	Switzerland Design: quasi-experiment Age: 8-11 N: 189 (EG=81; CG=108)	Experimental group: Guided Inquiry, teacher-directed structured practical work. Specific physics content included (Swiss MINT project). No direct teaching of control of variables strategy. Four curriculum units using guided inquiry to teach physics. Instructional guidance and scaffolding were a key design feature of the units. Text learning materials on floating/sinking, air/air pressure; spreading sound; stability of bridges. CPD on PCK of the relevant physics topics.	Focus: attainment, working scientifically: Measures included tests of physics content knowledge. Researcher-designed control of variables strategy test. Significant positive effect on conceptual physics understanding for the experimental group and significant positive effect on ability to apply the control of variables strategy correctly for the experimental

Study	Location and sample details	Description of approach	Reported outcomes
	Topic: physics	Control group: business as usual, i.e. direct teaching of general science only, as part of curriculum unit 'Human Beings and their Environment'.	group (d=0.75) and control group (d=0.50). A significant positive regression weight for the intervention variable (b=0.09, $p<0.05$. $R^2=0.40$) confirms the improved performance of students in the intervention classes on the CVS post-test compared to students in the control classes.
Zhang (2018)	USA Design: cluster RCT (not specified) Age: 9-11 N: 136 Topic: light	Experimental group: hands-on inquiry-based approach with or without withholding answers (hands-on + withholding answers=guided inquiry; hands-on only=structured inquiry or confirmatory inquiry) Control group: No hands-on strategy (direct teaching) Sample rather artificial since all students scoring >50% in a content pre-test are removed, all ELLs and pupils with SEN are also removed.	Focus: attainment, critical thinking Pre-and post-test to measure knowledge developed by researchers. Students in the 'hands-on only' group (structured inquiry) show significantly better scientific reasoning than students in other groups and marginally but not significantly better content knowledge than students in the other groups. No differences between groups on ability to apply science to real life situations.

Appendix 8: Summary quality assessments

Note: in the review text, strong corresponds to high quality, moderate to moderate quality and weak to low quality based on answers to the questions presented in Appendix 6.

B1 Assessment and feedback in primary science teaching

Author	**Cluster**	**Title**	**Global Rating**
Decristan *et al.* (2015a)	Assessment	Embedded Formative Assessment and Classroom Process Quality: How Do They Interact in Promoting Science Understanding?	Moderate
Decristan *et al.* (2015b)	Assessment	Impact of Additional Guidance in Science Education on Primary Students' Conceptual Understanding	Moderate
Ferrell *et al.* (2017)	Assessment	Audible Image Description as an Accommodation in Statewide Assessments for Students with Visual and Print Disabilities	Weak
Hondrich *et al.* (2018)	Assessment	Formative assessment and intrinsic motivation: The mediating role of perceived competence	Moderate
Hwang *et al.* (2018)	Assessment	Creating Interactive E-Books through Learning by Design: The Impacts of Guided Peer-Feedback on Students' Learning Achievements and Project Outcomes in Science Courses	Moderate
Hwang *et al.* (2021)	Assessment	Facilitating knowledge construction in mobile learning contexts: A bi-directional peer-assessment approach	Moderate

B2 Context-based and cross-curricular approaches to primary science teaching

Author	**Cluster**	**Title**	**Global Rating**
Burt *et al.* (2022)	Context-based and cross curricular approaches	New York City fourth graders who receive a climate change curriculum with hydroponic gardening have higher science achievement scores	Moderate
Fasasi (2017)	Context-based and cross curricular approaches	Effects of ethnoscience instruction, school location, and parental educational status on learners' attitude towards science	Moderate
Hardiman *et al.* (2017)	Context-based and cross curricular approaches	The effects of arts-integrated instruction on memory for science content	Moderate
Olgun & Adali (2008)	Context-based and cross curricular approaches	Teaching Grade 5 Life Science with a Case Study Approach	Weak
Qiao & Zhou (2020)	Context-based and cross curricular approaches	Research on the Integration of STEM Education into the Rural Elementary School Science Curriculum: An Example from Rural Elementary Schools in Western China	Weak
Zhang & Campbell (2012)	Context-based and cross curricular approaches	An Exploration of the Potential Impact of the Integrated Experiential Learning Curriculum in Beijing, China	Moderate

B3 Co-operative and collaborative approaches

Author	**Cluster**	**Title**	**Global Rating**
Chang & Hsin (2021)	Co-operative and collaborative approaches	The effect of the Self-explain–Discuss–Re-explain (SDR) learning strategy on high- and low-achieving fifth-grade students' achievement in science	Moderate
Chen, Hand & McDowell (2013)	Co-operative and collaborative approaches	The Effects of Writing-to-Learn Activities on Elementary Students' Conceptual Understanding: Learning About Force and Motion Through Writing to Older Peers	Moderate
Eysink *et al.* (2017)	Co-operative and collaborative approaches	Supporting primary school teachers in differentiating in the regular classroom	Weak
Hand *et al.* (2018)	Co-operative and collaborative approaches	Improving critical thinking growth for disadvantaged groups within elementary school science: A randomized controlled trial using the Science Writing Heuristic approach	Strong
Looi, Chen & Ng (2010)	Co-operative and collaborative approaches	Collaborative activities enabled by GroupScribbles (GS): An exploratory study of learning effectiveness	Moderate
Reeves *et al.* (2013)	Co-operative and collaborative approaches	Structural Equation Modeling of Knowledge Content Improvement using Inquiry Based Instruction	Moderate

B4 Critical thinking and argumentation in primary science

Author	**Cluster**	**Title**	**Global Rating**
Arias *et al.* (2017)	Critical thinking and argumentation	Justifying Predictions: Connecting Use of Educative Curriculum Materials to Students' Engagement in Science Argumentation	Moderate
Chen *et al.* (2016)	Critical thinking and argumentation	Using a modified argument-driven inquiry to promote elementary school students' engagement in learning science and argumentation	Moderate
Kara & Kingir (2021)	Critical thinking and argumentation	Implementation of the Model-Based Science Writing Heuristic Approach in Elementary School Science	Moderate
Miller *et al.* (2014)	Critical thinking and argumentation	The effects of reading to prepare for argumentative discussion on cognitive engagement and conceptual growth	Moderate
Sternberg *et al.* (2014)	Critical thinking and argumentation	Testing the Theory of Successful Intelligence in Teaching Grade 4 Language Arts, Mathematics, and Science	Strong
Tsai *et al.* (2012)	Critical thinking and argumentation	Using the Cognitive Apprenticeship Web-based Argumentation System to Improve Argumentation Instruction	Moderate

B5 Explicit instruction and related approaches in primary science

Author	Cluster	Title	Global Rating
Baumfalk *et al.* (2019)	Explicit instruction and related approaches	Impact of model-based science curriculum and instruction on elementary students' explanations for the hydrosphere	Moderate
Berry, Potter & Hollas (2013)	Explicit instruction and related approaches	Concept Maps and Informational Read-Alouds: Strengthening both Science and Reading for Elementary Students	Weak
Cohen & Johnson (2012)	Explicit instruction and related approaches	Improving the acquisition and retention of science material by fifth grade students through the use of imagery interventions	Moderate
Doabler *et al.* (2021)	Explicit instruction and related approaches	Efficacy of a Second-Grade Science Program: Increasing Science Outcomes for All Students	Strong
Kim *et al.* (2012)	Explicit instruction and related approaches	Project Clarion: Three Years of Science Instruction in Title I Schools among K-Third Grade Students	Weak
Michalsky, Mevarech & Haibi (2009)	Explicit instruction and related approaches	Elementary School Children Reading Scientific Texts: Effects of Metacognitive Instruction	Moderate
Randler (2009)	Explicit instruction and related approaches	Learning About Bird Species on the Primary Level	Moderate
Rotgans & Schmidt (2017)	Explicit instruction and related approaches	Interest development: Arousing situational interest affects the growth trajectory of individual interest	Moderate
Upadhyay & DeFranco (2008)	Explicit instruction and related approaches	Elementary students' retention of environmental science knowledge: connected science instruction versus direct instruction	Weak
van der Graaf *et al.* (2019)	Explicit instruction and related approaches	A combined approach to strengthen children's scientific thinking: direct instruction on scientific reasoning and training of teacher's verbal support	Moderate
Williams *et al.* (2009)	Explicit instruction and related approaches	Embedding Reading Comprehension Training in Content-Area Instruction	Moderate
Yeo *et al.* (2020)	Explicit instruction and related approaches	The Efficacy of an Image-to-Writing Approach to Learning Abstract Scientific Concepts: Temperature and Heat	Weak

Author	Cluster	Title	Global Rating
Zangori, Forbes & Schwarz (2015)	Explicit instruction and related approaches	Exploring the Effect of Embedded Scaffolding Within Curricular Tasks on Third-Grade Students' Model-Based Explanations about Hydrologic Cycling	Weak
Zheng *et al.* (2008)	Explicit instruction and related approaches	Effects of multimedia and schema induced analogical reasoning on science learning	Weak

B6 ICT supported and online teaching and learning in primary science

Author	Cluster	Title	Global Rating
Barak & Dori (2011)	ICT and online approaches	Science Education in Primary Schools: Is an Animation Worth a Thousand Pictures?	Moderate
Hodges *et al.* (2020)	ICT and online approaches	A quasi-experimental study comparing learning gains associated with serious educational gameplay and hands-on science in elementary classrooms	Moderate
Hu *et al.* (2019)	ICT and online approaches	Effects of question stem on pupils' online questioning, science learning, and critical thinking	Moderate
Hwang *et al.* (2020)	ICT and online approaches	Powering up flipped learning: An online learning environment with a concept map-guided problem-posing strategy	Moderate
Kim & Olaciregui (2008)	ICT and online approaches	The effects of a concept map-based information display in an electronic portfolio system on information processing and retention in a fifth-grade science class covering the Earth's atmosphere	Moderate
Looi *et al.* (2011)	ICT and online approaches	1:1 mobile inquiry learning experience for primary science students: a study of learning effectiveness	Weak
Sun *et al.* (2010)	ICT and online approaches	A 3-D virtual reality model of the Sun and the Moon for e-learning at elementary schools	Moderate
Wang & Tseng (2018)	ICT and online approaches	The Comparative Effectiveness of Physical, Virtual, and Virtual-Physical Manipulatives on Third-Grade Students' Science Achievement and Conceptual Understanding of Evaporation and Condensation	Moderate
Ward *et al.* (2013)	ICT and online approaches	My Science Tutor: A Conversational Multimedia Virtual Tutor	Moderate
Zacharia *et al.* (2016)	ICT and online approaches	The use of mobile devices as means of data collection in supporting elementary school students' conceptual understanding about plants	Moderate

B7 Language, literacy and text-based studies in primary science

Author	**Cluster**	**Title**	**Global Rating**
Alexander (2018)	Language, literacy and text-based approaches	Developing dialogic teaching: genesis, process, trial	Strong
Bigozzi *et al.* (2011)	Language, literacy and text-based approaches	The role of individual writing in fostering scientific conceptualization	Weak
Bravo & Cervetti, (2014)	Language, literacy and text-based approaches	Attending to the Language and Literacy Needs of English Learners in Science	Moderate
Cervetti *et al.* (2012)	Language, literacy and text-based approaches	The Impact of an Integrated Approach to Science and Literacy in Elementary School Classrooms	Strong
Cheng *et al.* (2015)	Language, literacy and text-based approaches	Learning effects of a science textbook designed with adapted cognitive process principles on grade 5 students	Moderate
Connor *et al.* (2017)	Language, literacy and text-based approaches	Acquiring Science and Social Studies Knowledge in Kindergarten Through Fourth Grade: Conceptualization, Design, Implementation, and Efficacy Testing of Content-Area Literacy Instruction (CALI)	Strong
Goldschmidt & Jung (2011)	Language, literacy and text-based approaches	Evaluation of Seeds of Science/Roots of Reading	Strong
Hanley *et al.* (2015)	Language, literacy and text-based approaches	Thinking, Doing, Talking Science Evaluation report and Executive summary	Strong
Hanley *et al.* (2020)	Language, literacy and text-based approaches	Thinking, doing, talking science: the effect on attainment and attitudes of a professional development programme to provide cognitively challenging primary science lessons	Strong
Henrichs & Lesemann (2014)	Language, literacy and text-based approaches	Early Science Instruction and Academic Language Development Can Go Hand in Hand. The Promising Effects of a Low-Intensity Teacher-Focused Intervention	Moderate

Jay *et al.* (2017)	Language, literacy and text-based approaches	Dialogic Teaching Evaluation report and executive summary	Strong
Kim *et al.* (2021)	Language, literacy and text-based approaches	Improving Elementary Grade Students' Science and Social Studies Vocabulary Knowledge Depth, Reading Comprehension, and Argumentative Writing: a Conceptual Replication	Strong
Kitmitto *et al.* (2018)	Language, literacy and text-based approaches	Thinking, Doing, Talking Science Evaluation report and executive summary	Strong
Lai & Chan (2020)	Language, literacy and text-based approaches	Enhancing Science Learning through Science Trade Book Reading for 5th Graders	Weak
Llosa *et al.* (2016)	Language, literacy and text-based approaches	Impact of a Large-Scale Science Intervention Focused on English Language Learners	Strong
Maerten-Rivera *et al.* (2016)	Language, literacy and text-based approaches	Effect of a multiyear intervention on science achievement of all students including English language learners	Strong
Vitale & Romance (2012)	Language, literacy and text-based approaches	Using in-depth science instruction to accelerate student achievement in science and reading comprehension in grades 1 -2	Moderate
Zwiep & Straits (2013)	Language, literacy and text-based approaches	Inquiry Science: The Gateway to English Language Proficiency	Moderate

B8 Learning outside the classroom in primary science

Author	**Cluster**	**Title**	**Global Rating**
Glick & Samarapungavan (2008)	Learning outside the classroom	Wolves are Beautiful and Proud Science Learning from a School Field Trip	Weak
Mills & Katzman (2015)	Learning outside the classroom	Examining the effects of field trips on science identity	Moderate
Piila *et al.* (2021)	Learning outside the classroom	STEAM-Learning to Mars: Students’ Ideas of Space Research	Weak
Scott & Boyd (2016)	Learning outside the classroom	Getting more from getting out: increasing achievement in literacy and science through ecological fieldwork	Moderate
Wells *et al.* (2015)	Learning outside the classroom	The Effects of School Gardens on Children's Science Knowledge: A randomized controlled trial of low-income elementary schools	Strong
Wünschmann *et al.* (2017)	Learning outside the classroom	Learning Achievement and Motivation in an Out-of-School Setting—Visiting Amphibians and Reptiles in a Zoo Is More Effective than a Lesson at School	Moderate

B9 Practical work in primary science

Author	Cluster	Title	Global Rating
Cvjetićanin *et al.* (2015)	Practical work, inquiry and investigation	The efficiency of student-led and demonstration experiments in initial physics-chemistry education in primary school	Weak
Dankenbring & Capobianco (2016)	Practical work, inquiry and investigation	Examining Elementary School Students' Mental Models of Sun-Earth Relationships as a Result of Engaging in Engineering Design	Moderate
Durmus & Bayraktar (2010)	Practical work, inquiry and investigation	Effects of Conceptual Change Texts and Laboratory Experiments on Fourth Grade Students' Understanding of Matter and Change Concepts	Moderate
Leuchter *et al.* (2014)	Practical work, inquiry and investigation	Designing Science Learning in the First Years of Schooling. An intervention study with sequenced learning material on the topic of 'floating and sinking'	Moderate
Meyer (2016)	Practical work, inquiry and investigation	The Benefits of Mouse Keeping—an Empirical Study on Students' Flow and Intrinsic Motivation in Biology Lessons	Moderate
Ünal & Aral (2014)	Practical work, inquiry and investigation	An Investigation on the Effects of Experiment Based Education Program on Six Years Olds' Problem Solving Skills	Weak

B9 Inquiry and investigation in primary science

Author	Cluster	Title	Global Rating
Chen & She (2015)	Practical work, inquiry and investigation	The Effectiveness of Scientific Inquiry with/without integration of scientific reasoning	Moderate
Chen *et al.* (2018)	Practical work, inquiry and investigation	Bridging the Gender Gap of Children's Engagement in Learning Science and Argumentation Through a Modified Argument-Driven Inquiry	Moderate
Di Mauro & Furman (2016)	Practical work, inquiry and investigation	Impact of an inquiry unit on grade 4 students' science learning	Moderate
Lai (2016)	Practical work, inquiry and investigation	Third Graders' Understanding of Air Concepts Facilitated by the iPod Inquiry Teaching Method	Weak
Lai *et al.* (2018)	Practical work, inquiry and investigation	The effects of computer-supported self-regulation in science inquiry on learning outcomes, learning processes, and self-efficacy	Moderate
Li *et al.* (2016)	Practical work, inquiry and investigation	The Effect on Pupils' Science Performance and Problem-Solving Ability through Lego: An Engineering Design-based Modeling Approach	Weak
Lin *et al.* (2009)	Practical work, inquiry and investigation	The Interplay of the Classroom Learning Environment and Inquiry-based Activities	Moderate
Letina (2016)	Practical work, inquiry and investigation	Effectiveness of Inquiry-Based Science and Social Studies Teaching in the Development of Students' Scientific Competence	Weak
Letina (2020)	Practical work, inquiry and investigation	Development of Students' Learning to Learn Competence in Primary Science	Weak
Polikoff *et al.* (2018)	Practical work, inquiry and investigation	The Impact of Speedometry on Student Knowledge, Interest, and Emotions	Strong
Schalk *et al.* (2019)	Practical work, inquiry and investigation	Improved application of the control-of-variables strategy as a collateral benefit of inquiry-based physics education in elementary school	Moderate

Author	Cluster	Title	Global Rating
Zhang (2018)	Practical work, inquiry and investigation	Withholding answers during hands-on scientific investigations? Comparing effects on developing students' scientific knowledge, reasoning, and application	Weak

Appendix 9: Refinements of methods from protocol

Searching

Google Scholar and JSTOR were not included in the search because of their limitations on the number of characters in search terms (150 and 200 respectively).

Codes for screening

An additional code 'no intervention' was added to the coding tool. This excluded studies where there was no stated intervention or approach. This is because the review aims to answer questions about the effectiveness of different approaches to primary science, so studies with no intervention do not allow us to answer the research questions.

Screening

Given the size of the sample of studies, it was not feasible within the time available to conduct the review to double screen 15% of all the studies, many of which were studies in health, medical and physical education and language education.

Instead, a pre-screening phase was introduced where all members of the review team screened the same 31 studies. An initial inter-screener reliability of 94% was calculated. The review team met first in pairs, then as a whole team to discuss disagreements and interpretations of the criteria, until there was 100% agreement on the 31 studies.

The first phase of screening was on the basis of title and abstract. The 10,458 studies were distributed amongst the review team. Any studies coded 'INCLUDE for second opinion' were discussed with one lead screener. This approach allowed decisions to be made efficiently and for less clear-cut decisions to be discussed with at least one other reviewer.

A total of 127 of 10,458 abstracts were screened in this way on title and abstract, and a further 63 out of 420 were screened on full text. All studies marked for inclusion were double screened.

Criteria for inclusion in evidence map

To ensure the review was feasible in the time and budget available, the inclusion criteria were amended to include studies which were published and included data collected between January 2007 and September 2021.

This document is available for download at https://educationendowmentfoundation.org.uk

The Education Endowment Foundation
5th Floor, Millbank Tower
21–24 Millbank
London
SW1P 4QP

https://educationendowmentfoundation.org.uk

@EducEndowFoundn

Facebook.com/EducEndowFoundn