Education Endowment Foundation

A review of evidence on primary science teaching

Review Protocol

November 2021

Lynda Dunlop, Judith Bennett, Pam Hanley, Michael Reiss and Maria Turkenburg-van Diepen

Table of contents

Introduction to the systematic review and protocol

The systematic review on primary science teaching will review approaches to teaching primary science education in order to identify which approaches, in which contexts and in which ways, are most effective. The review will comprise four strands of work:

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

The systematic review will be informed by two work packages: research on practice (to inform the decision on outcomes of interest, and to provide contextual information to ensure that the review is presented in a way that is relevant to teachers' needs) and a theoretical review (to explain likely mechanisms by which more effective approaches work).

It should be noted that some methodological decisions will be informed by the research on practice, theoretical review, the evidence map, and subsequent discussions with the EEF guidance panel (a group of experts in primary science teaching including – but not limited to – teachers, academics, representatives of subject associations and trusts). Once these exploratory strands of work have been completed, an addendum will be made to this protocol to ensure all aspects of the systematic review have been pre-specified prior to commencing the review. At present, this protocol offers explanations on how decisions will be made and how the findings from these strands of work will be utilised in the systematic review. Key decision-making points are identified, and at each of these points the protocol will be updated:

- November 2021: protocol updated with an appendix identifying proposed outcomes of interest for the review following the research on practice.
- December 2021: appendix proposing the format of evidence map to be submitted to the EEF.
- January 2022: appendix containing data extraction form to be submitted to EEF and details of methods for meta-analysis (if needed).

'Primary science education' is used throughout the protocol to refer to teaching and learning associated with the curriculum subject 'Science' at Key Stages 1 and 2 (ages 5-11). The early years are excluded from the review because the early years foundation stage (EYFS) is often subject to different aims, priorities, expected approaches, and areas of learning and the early years foundation stage therefore deserves separate attention. It is likely, given the different starting age of primary education in different countries, that some studies will include children under the age of 5. Such studies will be included in the review only where they also include children of the age group of focus (5-11). We recommend a separate review to look at the evidence on science in the early years because it would not be feasible to consider the EYFS within the scope of this review.

'Effectiveness' will be defined in relation to attitudinal and attainment outcomes, as well as others deemed important to teachers during the review of practice. The theoretical review will be used to explain the likely mechanisms by which approaches identified to be most effective work and identify useful explanations based on theory which support understanding of pedagogical approaches detailed within the guidance report recommendations.

Background and review rationale

Society depends on science to respond to pressing problems, from the Covid-19 pandemic to the climate crisis, so it is important to ensure the foundations for critical scientific literacy are laid in the primary school. Children should be provided with a science education that enables them to appreciate what science is, how it works, the social nature of its practices, and its relevance to the lives of individuals and society. This requires subject leaders and teachers to design and use effective curricula and pedagogies to develop substantive and disciplinary knowledge (Ofsted, 2021). Harlen (2015) argues that science should be focused on the development of a number of 'big ideas' of science, and about science. These fourteen big ideas, based on discussions with international science and science education experts, are culturally significant, have explanatory power, aid understanding, and lead to enjoyment and satisfaction. These are not the basis of the current national curriculum for England which applies to all local authority-maintained schools (Department for Education, 2014) – academies and free schools are able to determine their own broad and balanced curriculum.

Despite being a core subject in primary education in England, there is concern expressed about the amount of science taught in many primary schools, and the observed drop in pupils' performance in biennial national sample tests (Ofsted, 2021). Many children experience primary school science as consisting of fun activities rather than enabling deep learning (Bianchi, Whittaker and Poole, 2021), and few (5%) primary teachers hold specialist science degrees and teaching qualifications in these subjects (Royal Society, 2014). 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).

Primary science has been the focus of recent reviews and reports commissioned by the Wellcome Trust and others. The question of how prevalent different teaching approaches are in primary science has been addressed through surveys of science leaders, teachers, and children (Wellcome Trust, 2017). This indicates that the most common methods for teaching primary science in UK schools are teaching 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. The survey indicates that science is taught for an average of 1.7 hours a week, usually by classroom teachers rather than a science specialist. Other research with children in this age group has found that gender, parental attitudes to science, attitudes to school science and self-concept in science are associated with aspirations in science - but although children are positive about science, this does not translate into aspirations to be a scientist (DeWitt and Archer, 2015; DeWitt, Archer and Osborne, 2014). A recent survey indicates that against a background of broad enjoyment of primary science, there is a drop off in interest and enjoyment towards the end of primary school (Wellcome Trust, 2019).

A number of systematic reviews of relevance to primary science education have been published over the past ten years. These tend to cover both primary and secondary phases (Hartmeyer, Stevenson, and Bentsen, 2018; Nunes et al., 2017; Potvin and Hasni, 2014) and come from the perspective of understanding subject-specific concepts (Lelliott and Rollnick, 2010), assessment approaches (Hartmeyer, Stevenson, and Bentsen, 2018) and specific teaching approaches (Huerta and Garza, 2019). Whilst a number of relevant reviews on science aspirations (DeWitt and Archer, 2015; DeWitt, Archer and Osborne, 2014) and the impact of broader contextual factors (Banerjee, 2016) exist, these tend not to be systematic reviews focused on primary science.

One systematic 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). A systematic review that is both subject-(science) specific and phase-(primary) specific is needed to:

- Result in conclusions about the effectiveness of different approaches, of any duration, in achieving outcomes that are desired through primary school science education which are directly relevant to primary science subject leaders and teachers and which relate to attainment, attitudes and other outcomes they value.
- Synthesise knowledge on the most effective approaches to both curriculum design and pedagogy in improving pupil outcomes as reported in the international literature – including the possibility of differential approaches and outcomes for pupils who have been disadvantaged (for example, because of special or additional learning needs or those associated with socio-economic background).
- Integrate theoretical perspectives on primary science learning - this has the potential to explain why specific approaches are successful (or not) in achieving defined outcomes.

To ensure maximum utility to teachers, the systematic review will be informed by a theoretical review and research on practice (to determine the educational outcomes of primary science important to teachers in England, to identify where practice is consistent or inconsistent with research evidence, and to identify barriers to implementation of review recommendations).

The EEF has recently published a guidance report on improving secondary science teaching (EEF, 2018), based on evidence from a review of socio-economic status and science learning in formal educational settings (Nunes et al., 2017). The primary review will be determined by what emerges from the systematic review of evidence on primary science teaching. Consistency between the primary and secondary reviews will be identified, taking into consideration their different focus.

The review of evidence on primary science teaching is co-funded by the EEF and the Wellcome Trust.

Objectives

The overarching research question we will address in this review is:

'What approaches are most effective to improve pupil outcomes in primary science, in what context, and how?'

'Approaches' include curriculum and pedagogical strategies which are either science-specific (such as practical work), or broader strategies used in a primary science teaching context (such as small-group work/cooperative learning, dialogic teaching, assessment, feedback, metacognition and self-regulation). All search strings will include 'science'.

'Effectiveness' will be defined in relation to attitudinal and attainment outcomes, as well as others deemed important to teachers during the research on practice. Effectiveness will be determined according to whether or not there is a difference in outcome between students who experience an approach in classroom conditions, compared to business-as-usual or specified alternatives. We anticipate that the review will be able to identify what approach might be used, in which circumstances, to achieve which outcome. Answering this research question will enable the team to make recommendations for the development of a guidance report aimed at providing practical guidance to teachers, science subject leads, and head teachers about the approaches they take to teaching primary science.

The review will focus on the primary years, i.e. key stages 1 and 2 (age 5-11). Internationally, education for this age group can be referred to as basic education, elementary or middle school education, kindergarten, prep or foundation. The review will be confined to studies which include children aged 5-11 (with studies including children aged 3-4 included only where they also include those aged 5-11). The early years will be excluded because of the rapid developmental changes that occur during these years and the very different contexts (including reception classes in school, nurseries, and childminders) and the aims and focus of education during the early years. Science is not explicitly referred to as a separate subject in the early years foundation stage framework for England. A separate review would be necessary to produce recommendations that are sensitive to the demands of those working in early years childcare and education in England.

The review will comprise four strands of work: a theoretical review, research on practice, the production of an evidence map, and the systematic review. Each strand will focus on answering one or more research questions which contribute to the overarching research question. These research questions are presented under the corresponding strand headings.

Methodology

There will be four strands of work:

1. A theoretical review
2. Research on practice
3. Evidence mapping
4. Systematic review

Theoretical review

An iterative approach to the theoretical review will be taken. The purpose of the review of theory is 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 questions to be answered by this review are:

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

An open approach will be taken to identifying possible theories – drawing on expertise within the team, internet searches and examination of well-regarded sources of advice to teachers of primary science. Given the broad scope of theories – behavioural, cognitivist, constructivist to name a few – it will be important to include the perspectives on which interventions and approaches are based. We are interested in theories, at whatever level they are held to operate, that are fit for purpose and will examine the strength of evidence for these theories in both UK and international literature. In terms of curriculum, we are interested particularly in the balance between content and working scientifically, the delivery of the curriculum, and cross-curricular teaching of science. We interpret 'effective' to mean those approaches where there is strong evidence that they have a positive impact on the pupil outcomes of interest (attainment, attitudes and others identified in the research on practice). Particular attention will be paid to theories that consider pupils from disadvantaged backgrounds.

Some of the theoretical models of learning of relevance to primary science teaching begin with the early years and some are very 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. However, other theoretical models of learning are more specific, or, at any rate, are given more emphasis in primary science teaching.

One such concerns 'pupil misconceptions'. At its best, this takes seriously the distance between what a pupil already knows and what the teacher intends the pupil to learn. However, at its worst, it leads to a view of learning that denigrates out-of-school knowledge, giving the impression that science is distinct from the everyday world of the child and that a child's mind needs, somehow, to be cleared of its ideas about, for example, motion or the passage of the seasons before the child can receive and then bank valid scientific knowledge about Newton's first law of motion or the position of the Earth on its axis as it circles the Sun.

Another example is provided by constructivism, a model of learning that is widely applicable but has been particularly influential in primary science. As with pupil misconceptions, it exists in a version that is helpful to pupil learning and one that it less so. At its best, it takes seriously the idea that all learners need to construct their own understanding – knowledge cannot effectively be transferred from teacher to learners without learners actively reconstructing what they are taught. At its worst, though, it holds that direct teaching is pointless, even wrong, and that all scientific knowledge needs to be built up by learners through direct experience of the natural world.

We therefore anticipate that the theoretical review will produce a critical overview of the rather large number of theoretical models (more modestly, these might be referred to as 'presumptions') of learning of particular reference to primary science teaching.

The theoretical review will be updated at the following time points:

1. On conclusion of the research on practice, in order to ensure that the theory underpinning achievement of outcomes important to teachers are included in the review.
2. During the systematic review - where studies including a mechanism for action grounded in theory will be flagged for inclusion in the theoretical review.

The theoretical review will summarise in an intelligible form how children learn best in primary science, allowing the EEF to indicate mechanisms by which effective approaches are successful in the guidance report.

Research on practice

The research on practice serves two purposes.

The first purpose is to inform the direction of the systematic review by working with key informants to identify outcomes of interest for primary science in England beyond attainment and attitudes. We anticipate other outcomes relating to understandings of how science works, and to scientific practices including different methods of investigation will be important.

The second purpose is to identify the contextual factors (including barriers and enablers) which influence primary science teaching in England. This will enable the review team and guidance panel to be sensitive to how the best international research evidence corresponds to existing practices in England.

It is important that this strand of work is carried out first in order to ensure the systematic review includes pupil level outcomes relevant to teachers. This is because there are no primary science statutory tests, and only maintained schools are required to comply with the national curriculum. In order to make the guidance relevant to science education across a range of school types, the outcomes important for teachers must first be defined. The research on practice will include (i) a literature review focusing on practice in England and (ii) focus groups with primary teachers, subject leaders and headteachers. The literature review will summarise recent research on practice, and the focus groups will allow us to ask questions pertinent to the review and probe reasons for primary science practices.

The following research questions will be addressed during 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? (to be addressed following the completion of the systematic and theoretical reviews when their findings are compared with practice).

The research on practice will be a narrative 'state of the art' review (Grant and Booth, 2009) to identify current practice in primary science teaching in England. Note: this state-of-the-art review is not the systematic review (see below). The research on practice will include comprehensive searching of literature published between October 2013 and August 2021 (from the introduction of the latest iteration of the national curriculum in England, to present) to identify approaches used in the teaching of primary science (age 5-11) in England and any discussion of the aims and outcomes of these approaches. This review will draw on research literature, and that published in practitioner-focused journals such as Primary Science (published by the Association for Science Education) as well as reports published by, amongst others, the Wellcome Trust, Primary Science Teaching Trust, Primary Science Quality Mark, Ofsted and the British Science Association. Primary science experts on the EEF guidance panel are invited to suggest research and reports for inclusion. This review will make no formal quality assessment, and will synthesise findings from both qualitative and quantitative studies according to type of approach. It serves to identify what is happening in schools in England to contextualise the outcomes of the systematic review (below).

In parallel (in order to ensure the review is completed on time), online focus groups will be held, one set (three iterations, running for an hour each on different days/time) for primary science teachers, and the other for leaders (primary science subject leaders and/or headteachers). The focus groups will explore teachers' and science subject leaders' priorities, what they see as the aims of primary science, their anticipated and intended outcomes for primary science education, differences for different groups of children (including those from disadvantaged backgrounds) and contextual barriers and enablers associated with primary science. Answers to these questions were not found in a preliminary review of practice literature in England. Questions have been reviewed by the EEF and guidance panel and amended in light of comments (number of questions has been increased, then reduced; responding to comments from previous studies has been removed in the interests of time). Focus group discussions will be semi-structured to ensure the areas the review team wishes to be discussed (appendices 1 and 2) are covered, while giving latitude to respondents to contribute to the discussion agenda. The focus groups will last approximately one hour.

Ethical approval for the focus groups has been granted from the University of York Department of Education Ethics Committee, and details are available on the qualtrics form used for participants to register for the focus groups. See Appendix 1 for the focus group guide for teachers and Appendix 2 for the focus group guide for science subject leaders and headteachers.

We aim to recruit 3-6 participants for each focus group so that all participants have time to contribute in the hour they commit. Teachers and leaders will have different focus groups so that they can be asked different questions relevant to their post (see appendices 1 and 2). Separate focus groups will also allow teachers to speak openly about practice in their schools, for example where teachers' or leaders' perspectives are at odds with practices or policies.

Participants will be recruited using an open call distributed via the networks of the research team with support from the EEF and guidance panel. In addition, teachers, headteachers and science subject leads will be purposively contacted to include those in different types of schools, including schools with above average numbers of children eligible for free school meals and with English as an additional language. This will be achieved by contacting maintained schools and academies with these characteristics (using data available from the Department for Education and EEF Families of Schools databases via the school email address). Whilst we cannot guarantee that teachers in those schools will attend, we can present the findings of focus group discussions with contextual data on schools included in the focus group discussion so that we know who has participated and decide whether further recruitment is necessary (noting that further targeted recruitment will have an impact on the timing of the findings). Teachers will be asked to provide information on their role, teaching experience, and prior qualifications and experience in science education.

Focus groups will be audio-recorded, transcribed and analysed. A primarily deductive approach to qualitative analysis will be used. In the first round of coding, data will be coded according to (i) outcomes desired (ii) approaches used (iii) barriers experienced (iv) experiences of COVID (v) leadership experience and (vi) use of evidence. In the case of outcomes, these will be analysed by type, applying the framework developed by James and Brown (2007) which describes seven types of learning outcome: attainment, understanding, cognitive and creative, using, higher-order learning, dispositions, and membership, inclusion, and self-worth. Approaches, barriers, experiences of COVID, leadership experience and use of evidence will be analysed inductively, using codes identified from the data. Data will be coded by two researchers independently. For the inductive analysis of (ii) – (vi), researchers will each create a list of codes and definitions from the data, which will then be combined following discussion and used to code the data. The analysis relating to outcomes will be reported in November 2021, with the analysis of leadership, approaches, barriers, experiences during the pandemic and use of evidence following in 2022.

A short interim report of the focus group research will be prepared and used as the basis for the decision about which outcomes to include in the systematic review, to be submitted to the EEF as an appendix in November 2021 prior to the evidence mapping and systematic review stages. A report on the focus group research and the literature research on practice will be prepared as the systematic review is ongoing. It is intended that these findings will be used to inform the subsequent guidance report, taking into account perceptions of and barriers and facilitators to teaching primary science.

Evidence map

The evidence map is the first phase of the systematic review, and will characterise all of the research on the topic from studies that meet the inclusion criteria. The second phase of the systematic review involves narrowing the focus, then evaluating the included papers for their quality and synthesising the evidence.

The evidence map will allow us to decide which outcomes (attainment, attitudes and others from the research on practice) and factors (curriculum, pedagogy, pupil, teacher and contextual) to focus on in the systematic review. The question to be addressed by the evidence map is:

- 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 will present population characteristics, nature of intervention, comparators, outcomes, timing and setting (PICOTS) (Samson and Schoelles, 2012), as well as study design, sample size and other details as agreed in consultation with the EEF guidance panel and identified during the research on practice.

Valid research on outcomes of primary science education interventions is likely to be reported primarily in the science education and psychology research literatures. Studies for inclusion will be identified primarily from bibliographic databases (ERIC, JSTOR, Proquest Dissertation and Theses, PsycInfo, SSCI and BEI), citation searches, and by asking members of the EEF guidance panel to identify any grey literature. Preliminary searches of the BEI suggest that there is a substantial body of research to map: a search of scholarly literature published between 2007 and 2021 containing '(primary OR elementary OR "middle school")' AND 'science' AND 'education' AND '(attain* OR performance OR assess*)' in the abstract field returns 697 results for the first outcome of interest (attainment). Similarly, a preliminary search of '(primary OR elementary OR "middle school") AND 'science' AND 'education' AND (attitude* OR engage*)' returns 713 articles. To overcome publication bias to some degree, we will include PhD theses and grey literature as well as peer-reviewed research literature.

To narrow the number of studies included in the map, inclusion criteria will be applied (see systematic review, below). We will constrain the number of studies by: (1) language and (2) outcome variables of interest.

- Language: At this stage, we will include in our search studies not written in English, to determine whether any or all of these should be included in the systematic review. For instance, it might become apparent that studies of a certain primary science approach are more prevalent in a particular, non-English speaking country and appear in journals published in languages other than English.
- Outcome variables: these will identify the evidence about the effect of approaches to primary science curriculum and pedagogy on pupil level outcomes including attainment, attitudes and other relevant outcomes.

Evidence maps can be produced in a range of formats (Miake-Lye et al, 2016; Saran and White, 2018). The map will be presented in a format to be agreed between the review team and EEF when the key features of studies returned are apparent. The proposal for this, including a decision on whether or not we will use EPPI mapper or not, and why, will be presented in the appendix submitted to the EEF by December, 2021).

Inclusion and exclusion criteria for the evidence map

The inclusion and exclusion criteria outlined below, and grouped according to PICOS and SPIDER in Table 1.

Population/sample

The population is primary age/lower middle school age/elementary level school children, i.e. children aged 5-11 and teachers of children aged 5-11. Studies involving children younger than 5 will be included only when the sample also includes children age 5 or above. This is to ensure that findings are relatable to the English context for teaching science at key stages 1 and 2.

Interventions/practices/phenomena of interest

The review will focus on approaches to the teaching of science to children aged 5-11. Approaches can 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 will be included where they relate to science-specific contexts and/or outcomes.
Studies will be included in the review, subject to the exclusion criteria below, if they address the overall review research question: *What approaches are most effective to improve pupil outcomes in primary science, in what context, and how?* Studies included in the review will need to meet the criteria below:

1. The study addresses the review question (note that 'pupil outcomes' will be defined in the November 2021 appendix following the research on practice);
2. The study sample *includes* pupils aged 5-11 or teachers of children aged 5-11. Studies involving children aged 3 and above will be included only where they also include children of age 5 and above;
3. The study is undertaken in a formal educational establishment, most probably a school (including primary schools and middle schools and any other school for children aged 5-11). This is to ensure ecological validity and/or relatability of findings;
4. A principal focus on the study is the teaching of science (including biology, chemistry, physics and Earth science);
5. The approaches reported in the study are clearly defined approaches to teaching, curriculum design or assessment of primary science including, but not limited to intervention studies;
6. The study is published between January 2007 and September 2021 (to focus on studies published since the most recent curriculum reforms in the UK, and to ensure the project can be completed in time and within budget);
7. The study reports pupil outcomes (determined in advance by the review team and EEF in the appendix to be submitted in November 2021 following the research on practice focus groups) in relation to one or more of (a) pupils' knowledge and understanding of science, (b) pupils' knowledge of the nature of science or working scientifically, (c) inquiry skills (d) pupil attainment in science, (e) pupils' development of practical skills associated with science, and (f) pupils' affective responses to science and any other outcomes identified following the research on practice stage, to be agreed between the review team and the EEF in October 2021.

Studies will therefore be excluded from the review on the basis of the following criteria:

1. Do not include pupils aged 5-11, or teachers of pupils aged 5-11.
2. Does not take place in a formal educational setting;
3. Do not focus principally on the teaching of science (including biology, chemistry, physics and Earth Science);
4. Published before 2007 or after September 2021;
5. Does not report on the outcomes in Criterion 9 above.

PICOS	SPIDER	Scope	Explanation
Population	Sample	Children of primary age (5-11).	Studies involving children younger or older than 5-11 will only be included where the target age range is also included in the majority of the sample.
Intervention	Phenomenon of interest	Educational approaches to curriculum and pedagogy. Studies take place in educational settings.	Approaches can 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 will be included where they relate to science-specific contexts and/or outcomes. The review will focus 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 will be excluded.
Comparison	Design	Studies should include a counterfactual.	Process evaluations of randomised control trials (RCTs) which primarily consist of qualitative and/or survey data will be considered where they either meet the inclusion criteria above, or where the RCT they refer to meets the inclusion criteria. For qualitative studies, the hypothesised counterfactual should be explicit, plausible or identified in the data.
Outcome(s)	Evaluation	To be determined: see addendum 1.	Studies reporting attainment, attitude and the other outcomes of interest identified in addendum 1 will be included. All other outcomes will be excluded.
Study design	Research type	All study designs	All study designs which allow us to answer the review questions will be included.

Table 1: inclusion and exclusion criteria classified using PICOS and SPIDER (Methley et al., 2014). Further refinement of inclusion and exclusion criteria will follow the review of practice.

Search strategy for identification of studies

The focus of the search will be articles, reports, theses and other publications on approaches to teaching primary science. Studies will be identified from the following sources:

- Bibliographic databases: Australian Education Index, ERIC, JSTOR, PsycInfo, SSCI and BEI

- Google Scholar
- Citation searches of key authors/papers
- Reference lists of key authors/papers
- References on websites of organisations used by teachers as sources of evidence to inform teaching (as identified in the research on practice, therefore cannot currently be predetermined.)
- Direct requests to personal contacts
- Direct requests to key organisations
- Focus groups with key informants
- Proquest database of dissertation and theses
- EEF database

Search period

November 2021 - January 2022 (subject to contract, review of protocol and appointment to research posts)

Search terms

Indicative search terms (to be included in title, abstract or keywords) are found below. Note:

1. Where 'primary' is used as a search term, the string 'OR middle OR elementary OR 'early childhood' OR kindergar*' will be added.
2. Where 'scien* is used a a search term, the string 'OR biolog* OR chemi* OR physic*' will be added.

Primary and science will be used in all searches. It would not be feasible to identify and search for every concept or topic included in primary science.

A preliminary search in one database for primary AND scien* AND 'school OR education' in the abstract field returns >600 articles published since 2007. We therefore propose to search in the following phases:

By outcomes of primary science

primary AND scien* AND learning
primary AND scien* AND 'attain* OR assess* OR outcome OR test OR grad* OR prog* OR feedback'
primary AND scien* AND understand*
primary AND scien* AND 'misconception OR preconception OR "alternative understand*"
primary AND scien* AND attitud* OR engage
primary AND scien* AND socio-econ* OR income OR SES OR disadvantag*
primary AND scien* AND represent* OR participat* OR capital

Note: this review will be informed by the practice review. Any additional search terms resulting from outcomes will be identified in the November 2021 addendum to the protocol.

Studies from this set of searches will be included where they meet the inclusion criteria (note the criterion about reporting clearly defined approaches to teaching, curriculum design or assessment of primary science).

By approaches to primary science

- primary AND scien* AND curriculum
- primary AND scien* AND 'teach* OR pedagog*'
- primary AND scien* AND practical OR investigat* OR skill OR inquir* OR enquir* or question*
- primary AND scien* AND 'collaborat* OR group OR discussion OR talk OR dialog*'
- primary AND scien* AND 'thinking* OR reason*'
- primary AND scien* AND 'vocabulary OR language OR linguistic'
- primary AND scien* AND 'metacogniti* OR self-regulat* OR memor* OR "cognitive strateg*"'

Search terms will be updated on conclusion of the research on practice. If additional outcomes or specific approaches are identified, these will be included in the appendix to the protocol submitted in November 2021.

Use of EPPI-Reviewer 4

EPPI-Reviewer 4 will be used to facilitate the storage, organisation and management of documents, data extraction and data analysis. All the records of the research studies identified will be uploaded onto the online tool for systematic reviews. We do not intend to use the text mining function built in to EPPI reviewer. The EPPI online support states that this can be very useful for larger reviews (50,000 references or more) than the one commissioned.

Outcomes

The principal outcome will be an evidence map (or overview) of the research into approaches that have an impact on pupils' cognitive and affective outcomes in primary science. The precise form of the evidence map will be agreed with the EEF and the guidance panel for the Improving primary science guidance report. The intention would be to make the map as user-friendly as possible. It does not make sense to pre-specify the format of this report until it is clear what the review has found and how this relates to current practice.

Screening and selection of studies

After removing duplicates, studies will initially be screened on the basis of titles and abstracts. The criteria will be applied to a sample of the studies (the sample size will depend on the number that emerge) to check the efficacy of the search strategy. The full screening process will then take place, with 15% of the records being double-screened by two members of the team independently. If there is 95% agreement, the remaining records will be single-screened. Disagreements will be resolved through discussion, or through a third person adjudicator from within the review team if necessary. Any refinements will be added to the inclusion criteria if required. If agreement is lower than 95%, additional records will be double-screened in blocks of 50 until 95% agreement is obtained.

Full reports will be obtained for those studies that, on the basis of the abstract, appear to meet the criteria. Full reports will also be obtained if there is insufficient information to judge whether or not the report meets the criteria. The full reports will then be screened, with those meeting the criteria included in the evidence map. As with screening on the basis of abstracts, a 15% sample of the reports will be double-screened, with the target of achieving 95% agreement.

Where there are two or more reports of the same study, all versions will be obtained, and one designated as the 'key' report on the basis of detail provided and place of publication. The other reports will be noted as linked reports.

To mitigate publication bias as far as is feasible, the final list of studies (we anticipate the review including 100-150 studies) to be included will be circulated to relevant members of the EEF guidance panel. Members will be asked to identify any potential missing sources, including grey literature.

The results of this process will be documented in a PRISMA flow chart (Moher et al., 2009). We do not anticipate using machine automation during the review production. If this changes, it will be noted and details provided in the appendix to be submitted in January 2022.

Systematic review

The overall research question this review addresses is: What approaches, in which contexts and in which ways, are most effective to improve pupil outcomes in primary science? The review will be presented in accordance with the PRISMA 2020 checklist (Page et al., 2021). Research questions the systematic review will address are:

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

The findings will be limited to those approaches about which research evidence exists. For example, although teachers and subject leaders may want to know the relative pros and cons of inquiry-based lessons compared to teacher-led lessons, it may not be possible to make such a comparison if the evidence does not exist, or does not enable such comparisons to be made.

Note here that 'disadvantage' might refer to a number of circumstances. We will report any outcomes that are different for children with special educational needs and/or disabilities, those from low income backgrounds, who have English as an additional language and any other disadvantage reported in research studies. To maintain the focus on primary science, approaches to assessment, feedback, dialogic teaching, cognitive strategies, e/inquiry and pre/misconceptions will be included only when used in a primary science context, and where the outcomes of interest are science attainment, attitudes towards science or the other outcomes of interest agreed following the research on practice.

Moderating factors will be identified from research studies included in the review and may include those associated with children (e.g. sex, socio-economic status, prior attainment), teachers (e.g. qualifications, experience), or approaches (e.g. duration, support).

Depending on what emerges in the evidence map, there may be a need to refine the research question to identify a group of reports of a size that can be managed within the time available. Should this be the case, a decision-making heuristic will be created and submitted to the EEF in the January 2022 addendum to ensure that decisions are not made on the basis of preferences for particular outcome or approaches. This will be informed by the research on practice (which identifies approaches to primary science education in England). Appropriate inclusion/exclusion criteria will be drawn up should the review question be refined.

The population/sample and the Interventions/practices/phenomena of interest are as described for the evidence map. Studies identified as meeting the inclusion criteria for in depth review will be analysed in depth using EPPI-Reviewer 4. Independent double screening will be used to select studies for inclusion to avoid studies being missed, as long as the review is not delayed (single screening by an experienced reviewer can be appropriate where time to conduct the review is shortened).

Data extraction and management

Data will be extracted using EPPI-Reviewer 4 software with a bespoke coding framework based on its available schemas, the Cochrane risk of bias tool (Higgins et al., 2019) those in the EEF main data extraction tool and effect size data extraction tool (EEF, 2019) for quantitative studies, and the Standards for Reporting Qualitative Research (SRQR) (O'Brien et al., 2014) and Consolidated Criteria for REporting Qualitative (COREQ) research (Tong et al., 2007). Additional codes, and/or explanatory guidance for reviewers, will be added if necessary following familiarisation of the research team with EPPI Reviewer in late 2021. The data extraction form will be submitted to the EEF as an appendix in January 2021.

The extraction tool will be used to identify and record the chief characteristics of studies on primary science education, including study location, research questions, nature/duration/frequency of intervention, age of children, role of teachers or subject leaders, relationship between researcher(s) and intervention, and theoretical underpinning of the approach, as well as details of the sample, data collection and analysis. The data extraction form will be piloted by senior members of the project team on a subset of papers and refined following discussion. Sample fields to be included in the extraction form can be found in Appendix 3. The complete extraction forms are not included here as they extend to over 24 (EEF main data extraction tool) and 10 (Cochrane risk of bias tool) pages and are available in full elsewhere (EEF, 2019; Higgins et al., 2019) and will be refined, with the final full extraction tool determined in advance of the extraction phase (January 2022).

In the final review, key features of the extracted studies will be presented in tabular format, showing characteristics such as methodology, approach, sample composition and size, main results, quality appraisal. Tables will be accompanied by a narrative review, to include a contextual analysis and discuss moderating and mediating influences where the evidence is available.

Appraisal of included studies

Study quality will be determined with reference to both the operational and theoretical/conceptual definition of the outcome of interest, research and sampling design, sample, reliability and validity (or credibility, authenticity, relatability for qualitative studies) and data analysis. Appendices 4 and 5 list the features of quantitative and qualitative studies that will be used to appraise quality, and these will be supplemented by existing tools where appropriate (for example, for randomised controlled trials, this will be the revised Cochrane risk-of-bias tool (RoB 2) and where appropriate the test versions available for cluster and crossover designs. For non-randomised studies, ROBINS-I will be used). Should detail beyond that provided in the appendices be required in advance, it will be included in the appendix to be submitted in March 2022.

In the narrative synthesis we will examine the relationship between studies to give an overall assessment of the evidence on the effectiveness of teaching approaches on pupil outcomes in different primary science education contexts in order to respond to the research questions.

In judging the overall quality of the reports, three components contribute to the judgement on the weight of evidence in the reports. These are the:

- soundness of studies (internal methodological coherence), based upon the report only;
- appropriateness of the research design and analysis used for answering the review question; and
- relevance of the report topic focus (from the sample, measures, scenario, or other indicator of the focus of the study) to the review question.

Key considerations regardless of the study methodology will be sample size and composition, real-world replicability (eg affordability, who delivered the intervention in the study), duration and frequency of intervention, validity of outcome measure.

Notes on meta-analysis and synthesis

These can only be performed on the appropriate data. Methods will be determined when the search has been completed, and methods for calculating effect size, dealing with missing data, meta-analysis, investigation of heterogeneity and sensitivity analysis will be submitted to the EEF in March 2022. Where this will not be performed, the decision will be justified (the most likely reason being that there will be few, if any, studies of the same approach using similar measures).

Data synthesis

A narrative synthesis of the included studies will be used to answer the review questions. Where appropriate we will aim to do a meta-analysis (i.e. where studies are sufficiently similar in outcome measures). Effect sizes and associated measures of uncertainty will be calculated where studies present the necessary information to do so. The intervention outcomes of interest for this review are *attainment, attitudes*, and other outcomes to be identified in the review of practice (see addendum 1). For quantitative studies, these outcomes are likely to be reported as scores. It is unlikely that measures will be directly comparable, but where they are, they will be used to calculate standardised effect sizes. Should we find studies where there are comparable outcome measures, they will be used in a meta-analysis. Where studies report a range of outcomes for a single intervention (e.g. attainment and attitudes), we will, where the data permits, calculate an effect size for each outcome. For studies with outcome measures at different time points, the outcome immediately following the completion of the intervention will be used in the synthesis.

To answer review question on the moderating factors that influence approaches, and the barriers and facilitators to the use of more effective approaches, we will extract qualitative and quantitative information from the included studies, where available. We will use an adapted approach to thematic synthesis reported by Thomas and Harden (2008). This will involve coding text, developing descriptive themes and creating analytical themes to go beyond the findings of included studies to answer the review questions and ultimately to identify moderating factors, barriers and facilitators to effective primary science education.

To answer the review question about whether certain approaches to primary science teaching are more effective at improving attainment, attitudes or other outcomes determined in addendum 1, we will use regression analysis to compare intervention type for moderating effects.

The analysis will be supported by the data synthesis process, which will include a descriptive analysis to summarise the characteristics of approaches used in the included studies. The descriptive analysis will include a synthesis of sample size, attrition, types of approach, duration of interventions, educational settings, participant characteristics and effect size of *attainment, attitudes* and other outcomes identified in addendum 1.

Synthesis of findings

The data will be synthesised to bring together the studies which answer the review question and sub-questions, and which meet the quality criteria relating to appropriateness and methodology.

Outcomes

The principal outcome will be a report synthesizing the research evidence on the impact on pupil outcomes of approaches to teaching science at primary school level, with attention to how these vary by context and level of disadvantage (noting that the measures of disadvantage are likely to be different in different countries and contexts). The report will take the form of a narrative synthesis with tables and figures where appropriate. Other possible formats will be discussed with the EEF.

Limitations

The limitations of these review criteria include language bias. We will use search terms in English only. Whilst abstracts are often available in English, articles are not always translated into English. We will exclude articles in a language other than English.

Reporting

The primary output will be a publishable research report using the EEF review reporting template for evidence reviews. The research report will interpret the findings in response to the research questions. The systematic and theoretical reviews will reflect on the broader social and political context, situating conclusions in light of enabling and constraining factors associated with primary science education identified in the research on practice. We anticipate that the report will have separate chapters for each of the four strands of work (theoretical review, research on practice, evidence map and systematic review). If, upon completing the work, the review team considers it wise to present an alternative structure that is simple and easy to follow, this will be discussed in advance of the submission of the draft.

The review report will connect 'effective' approaches with theoretical underpinnings to explain why some approaches are more effective than others at achieving outcomes of interest, highlight any specific approaches useful for narrowing gaps in attainment/attitudes for different groups of pupils and interpret research evidence in relation to the research on practice and identify priority actions at different scale (e.g. class, school, policy). The review findings will be limited by the research which exists. Where the guidance panel, teachers or subject leaders have specified interest in specific approaches, and no evidence is found, absences will be identified in the report and reflected in the evidence map.

Findings will be related to existing reviews – both science specific (Nunes et al., 2017) and relevant non-science-specific reviews (e.g. Muijs and Bokhove, 2020), avoiding duplication and noting important points of overlap and contrast between primary and secondary science. We anticipate collaborating closely with the EEF to ensure we produce a review that is robust and relevant, and which is presented in a way which enables the findings to be most easily translated to practice. The EEF's evidence to decision framework will be used to assess the quality of evidence, considering factors such as priority, feasibility, cost, acceptability, equity, transferability, priority, certainty of evidence when developing recommendations.

The secondary output will be the publication of the evidence map. The evidence map will most likely be presented by type of intervention, age group and/or desired outcome, but this decision is most usefully made when the studies have been identified.

The review team will work with the EEF guidance panel to produce a school-facing guidance report using the findings from the review presented in a way to best inform practice.

Personnel

The review team brings together expertise in science education, including primary science practice and research and conducting systematic reviews (Bennett, 2008, Bennett et al, 2010; 2018; 2019;

2020; Collins et al, 2010; Dunlop et al, 2015; 2020; Rushton and Reiss, 2020) and evaluation of classroom interventions (Hanley et al., 2020). We draw together expertise from the University of York (University of York Science Education Group, based in the Department of Education, and York Trials Unit, based in Health Sciences) and University College London Institute of Education.

University of York

Dr Lynda Dunlop (Co-Principal Investigator). Senior Lecturer in Science Education. She has experience conducting systematic reviews and empirical research with primary science teachers and children. She was responsible for science strands on Primary PGCE, BEd and CPD programmes, and led a primary science project funded by the Primary Science Teaching Trust (PSTT). Lynda will be responsible for project management and deliverables.

Professor Judith Bennett (Co-Principal Investigator). Salters' Professor of Science Education and leader of the University of York Science Education group (UYSEG). She led the Science Review Group of the Evidence-Informed Policy and Practice Initiative (EPPI). She has significant experience in curriculum development and research, and currently holds grants that focus on primary science. Judith will be responsible for the strategic direction of the study.

Dr Pam Hanley. Research Fellow in the York Trials Unit. She has co-authored a systematic review of primary science programmes (NSF-funded) and led a rapid evidence assessment of policy and practice in practical science (Gatsby-funded). Her mixed methods research background, particularly in primary and secondary science, includes several EEF-funded evaluations. She is currently Co-PI of the Thinking, Doing, Talking Science (primary science approach) trial.

Dr Maria Turkenburg-van Diepen. Post-doctoral research associate. Maria is a very experienced researcher who has undertaken a wide range of studies with primary-aged children, including with the Centre for Industry Education Collaboration (CIEC). She has been involved in a number of systematic and rapid evidence reviews and will support strand leaders, primarily assisting with focus groups, evidence map and systematic review.

An additional post-doctoral researcher will be appointed to support strand leaders for the duration of the project and two doctoral researchers will support with the data extraction stage of the review. These posts will be advertised and appointed to coincide with the start of the systematic review.

University College London Institute of Education

Professor Michael Reiss. Professor of Science Education at UCL Institute of Education. The former Director of Education at the Royal Society and Chief Executive of Science Learning London, he spent six years running an ITE course for primary BEd teachers and has published widely on policy and practice in 5-19 science education. Michael will be responsible for the theoretical review.

Conflicts of interest

None of the review personnel have any conflicts of interest to declare.

Registration

The review protocol will be registered on the Open Science Framework registry following peer review. We anticipate publication of at least one paper in a peer-reviewed science education journal and at least one article in a practitioner-focused publication.

Timeline

Note: key points where the guidance panel will be asked for input and the protocol will be updated are highlighted in orange.

Dates	Activity	Staff responsible/ leading
July 2021	Ethics approved Guidance panel meeting	LD (completed)
August 2021	Protocol submitted	LD (completed)
September 2021	Protocol reviewed	EEF (completed)
September 2021	Research on practice: focus groups held	LD (completed)
November 2021	Research on practice completed	LD
November 2021	Outcomes of interest for systematic review determined	Review team and EEF
November 2021	Searching and Mapping	PH
December 2021	Format of evidence map determined	Review team and EEF
January 2022	Evidence map produced	PH
January 2022	Final version of data extraction form determined	Review team and EEF
March 2022	Decision on meta-analysis to be made. Where this is to be conducted, the review team will submit methods.	Review team
May 2022	Data extraction, synthesis and quality appraisal completed: systematic review completed	JMB
May 2022	Theoretical review completed	MJR
June 2022	Draft review report submitted to EEF and reviewed	EEF
July 2022	Final report submitted	LD
October 2022	Ongoing support: guidance report produced	EEF

References

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, DOI: 10.1080/2331186X.2016.1178441.

Bennett, J. (2008) The evidence on attitudes towards science and mathematics for 14-19 year-olds. In: The Royal Society State of the Nation report: science and mathematics education, 14-19. pp 171-179. London: The Royal Society.

Bennett, J. M., Airey, J. N., Dunlop, L., & Turkenburg, M. G. W. (2020). The impact of human spaceflight on young people's attitudes to STEM subjects. *Research in Science and Technological Education*, *38*(4), 417-438. https://doi.org/10.1080/02635143.2019.1642865.

Bennett, J. M., Dunlop, L., Knox, K. J., Reiss, M., & 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*(14), 1755-1773. https://doi.org/10.1080/09500693.2018.1511936.

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. https://doi.org/10.1080/09500693.2018.1563729.

Bennett, J., Hogarth, S., Lubben, F., Campbell, B., & Robinson, A. (2010). Talking science: The research evidence on the use of small group discussions in science teaching. *International Journal of Science Education*, *32*(1), 69-95. https://doi.org/10.1080/09500690802713507.

Bianchi, L., Whittaker, C. & Poole, A. (2021). *The ten key issues with children's learning in primary science in England.* https://seerih-innovations.org/media/doc/Childrens_Learning_in_Primary_Science_Report_2020.pdf.

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.

Department for Education. (2014). Statutory guidance. National curriculum in England: framework for key stages 1 to 4. https://www.gov.uk/government/publications/national-curriculum-in-england-framework-for-key-stages-1-to-4

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.

Dunlop, L., Compton, K., Clarke, L., & McKelvey-Martin, V. (2015). Child-led enquiry in primary science. *Education 3-13*, *43*(5), 1-22. https://doi.org/10.1080/03004279.2013.822013.

Dunlop, L., Airey, J. N., Bennett, J. M., & Turkenburg, M. G. W. (2020). Close encounters between young people and human spaceflight. *International Journal of Science Education Part B*. https://doi.org/10.1080/21548455.2020.1730018.

EEF (2019) EEF main data extraction v 1.0 October 2019 [Standard] https://educationendowmentfoundation.org.uk/public/files/Toolkit/EEF_main_data_extraction_v_1.0_October_2019.pdf.

EEF (2019) EEF Toolkit effect size data extraction v 1.0 October 2019 [Standard] https://educationendowmentfoundation.org.uk/public/files/Toolkit/EEF_Toolkit_effect_size_data_extraction_v_1.0_October_2019.pdf

Glover, J., Izzo, D., Odato, K., and Wang, L. (2006). EBM Pyramid and EBM Page Generator, Trustees of Dartmouth College and Yale University.

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.

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. (2015). Towards big ideas of science education. *School Science Review* 97(359), 97-107.

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, DOI: 10.1080/0969594X.2017.1377685.

Higgins, J.P.T., Savović, J., Page, M.J., & Sterne, J.A.C. (2019). Revised Cochrane risk-of-bias tool for randomised trials (RoB2). https://sites.google.com/site/riskofbiastool/welcome/rob-2-0-tool?authuser=0.

Hoover, W. A. and Gough, P. B. (1990) The Simple View of Reading, *Reading and Writing,* 2 (2), 127–160

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). *Educational Psychology Review*, *31*(3), 533-570. https://doi.org/10.1007/s10648-019-09477-1.

James, M., & Brown, S. (2005). 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

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.

Methley, A. M., Campbell, S., Chew-Graham, C., McNally, R., & Cheraghi-Sohi, S. (2014). PICO, PICOS and SPIDER: a comparison study of specificity and sensitivity in three search tools for qualitative systematic reviews. *BMC Health Services Research*, 14(1), 579. https://doi.org/10.1186/s12913-014-0579-0

Miake-Lye, I. M., Hempel, S., Shanman, R., & Shekelle, P. G. (2016). What is an evidence map? A systematic review of published evidence maps and their definitions, methods, and products. *Systematic reviews*, *5*(1), 1-21. DOI 10.1186/s13643-016-0204-x.

Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). *P*referred *R*eporting *I*tems for *S*ystematic Reviews and *M*eta-*A*nalyses: The PRISMA Statement. PLoS Med 6(7): e1000097.

Muijs, D., & Bokhove, C. (2020). *Metacognition and self-regulation: Evidence review.* Education Endowment Foundation. London. Available at: https://educationendowmentfoundation.org.uk/evidence-summaries/evidence-reviews/metacognition-and-self-regulation-review.

Mullis, I.V.S., Martin, M.O., Foy, P., Kelly, D.L., & Fishbein, B. (2020). TIMSS 2019 International Results in Mathematics and Science. Retrieved from Boston College, TIMSS & PIRLS International Study Center website: https://timssandpirls.bc.edu/timss2019/international-results/.

Nunes, T., Bryant, P., Strand, S., Hillier, J. M., Barros, R., & Miller-Friedmann, J. (2017). Review of SES and science learning in formal educational settings. https://educationendowmentfoundation.org.uk/public/files/Review_of_SES_and_Science_Learning_in_Formal_Educational_Settings.pdf.

O'Brien, Bridget C. PhD; Harris, Ilene B. PhD; Beckman, Thomas J. MD; Reed, Darcy A. MD, MPH; Cook, David A. MD, MHPE Standards for Reporting Qualitative Research, Academic Medicine: September 2014 - Volume 89 - Issue 9 - p 1245-1251 doi: 10.1097/ACM.0000000000000388

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

Page, M. J., Moher, D., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow et al. (2021). PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. *bmj*, *372*(71) doi:10.1136/bmj.n71.

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, DOI: 10.1080/03057267.2014.881626.

Rushton, E.A., & Reiss, M.J. (2020). 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*, 1-63. https://doi.org/10.1080/03057267.2020.1799621.

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/.

Saran, A., & White, H. (2018). Evidence and gap maps: a comparison of different approaches. *Campbell Systematic Reviews*, *14*(1), 1-38. https://doi.org/10.4073/cmdp.2018.2

Schünemann, H., Brożek, J., Guyatt, G. and Andrew Oxman, A. (2013). GRADE Handbook. Available at: https://gdt.gradepro.org/app/handbook/handbook.html.

Slavin, R.E., Lake, C., Hanley, P., & 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.libproxy.york.ac.uk/10.1002/tea.21139.

The Royal Society. (2014). Vision for science and mathematics education. https://royalsociety.org/-/media/education/policy/vision/reports/vision-full-report-20140625.pdf

Thomas, J., & Harden, A. (2008). Methods for the thematic synthesis of qualitative research in systematic reviews. *BMC medical research methodology*, *8*(1), 1-10.

Tong, A., Sainsbury, P., & Craig, J. (2007). Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. *International journal for quality in health care*, *19*(6), 349-357.

Wellcome Trust (2017). *State of the nation report of UK primary science education* https://cms.wellcome.org/sites/default/files/state-of-the-nation-report-of-uk-science-education.pdf.

Wellcome Trust (2019). *What pupils think of science in primary schools.* https://cms.wellcome.org/sites/default/files/what-pupils-think-of-science-in-primary-schools.pdf.

Appendix 1: Focus group guide (primary teachers)

The purpose of the focus group is to find out what is important for teachers and what currently happens in schools from year 1 to year 6. We will use what you tell us to help us decide the focus of our systematic review of primary science. We value your honesty.

Introductions

Please introduce yourself with information about your teaching experience, your current year group, role in school and role teaching science in your school.

Background information

- What do you see as the value of science (as a discipline) to society?
- What do you see as the value of science education to individuals?

Learning science

- What do you see as the main aims of school science in primary school?
- What do you think it is important for children to learn about in their science lessons by the end of their primary education?
- What are the most important things you would expect children to be able to do if you were to judge primary science education to be successful?
- How would you rate the following outcomes in terms of importance:
    - Positive attitudes towards science
    - Attainment in science
    - Other (please specify)
    - probes: what do you understand by attitude, attainment (and any other outcomes identified)

Teaching science

- Using examples from lessons you have taught, can you tell me:
    - What sort of approaches do you use (and why)?
    - Do you use different approaches with different groups of children (e.g. children with specific special educational needs)?
    - How much is for conceptual development, procedural skills and how much is understanding how science works? How did these support each other?
    - What do you find most challenging in teaching primary science?
- How do you resource your science lessons?
    - Do you use any published resources in your lessons? Which? How do you decide what to use?
    - What time is allocated to primary science in your year group (typical hours/week)
- Do you use any research or guidance to inform your practice?
    - If yes, what, and how did you find out about it?
- What challenges do you experience in teaching primary science?
- What sort of support do you think is needed for primary science in your school to thrive?

Assessment

- How do you assess children in science? Is this common across the school, or just in your class?
- How do you use your science assessment data?

Your curriculum

- How is science teaching organised in your school?
- Are you involved in planning your science curriculum?
    - If yes, in what ways? How do you decide what to prioritise over the short/medium/long term?

Closing questions

- How have things been different as a result of the pandemic - are there any aspects of primary science practice from the past 18 months that you will retain, and any that you are keen to say goodbye to?
- Have you used the EEF Secondary Science guidance report to inform your practice? Which aspects, and why?
- How do you feel about teaching primary science?
- What should be included in a good guidance report for primary science?
- Is there anything that we have not discussed that you think it is important that we include in our review?

Appendix 2: Focus group guide (primary subject leaders and headteachers)

The purpose of the focus group is to find out what is important for teachers and what currently happens in schools from year 1 to year 6. We will use what you tell us to help us decide the focus of our systematic review of primary science. We value your honesty!

Introductions

Please introduce yourself with information about your leadership role and your involvement in science teaching in your school.

Background information

- What do you see as the value of science (as a discipline) to society?
- What do you see as the value of science education to individuals?

Learning science

- What do you see as the main aims of school science in primary school?
- What do you think it is important for children to learn about in their science lessons by the end of their primary education?
- What are the most important things you would expect children to be able to do if you were to judge primary science education to be successful?
- How would you rate the following outcomes in terms of importance:
    - Positive attitudes towards science
    - Attainment in science
    - Other (please specify)
    - probes: what do you understand by attitude, attainment (and any other outcomes identified)

Your school curriculum

Thinking about the curriculum in your school

- What do you prioritise when creating your curriculum? Why?
- How would you describe the proportion of the curriculum focused on conceptual development, knowledge about the nature of science, procedural skills and inquiry methods?
- What difficulties or challenges do you face in planning your science curriculum?
- What areas of your curriculum do you think need further development? Why?
- Does your school make use of evidence to inform Science teaching and learning? What evidence? Where do you find out about it? How is it used?

Assessment

- Do you have a whole school approach to assessment in science? If yes, why, and what is this?
- What do you assess (probe for: conceptual knowledge, problem solving, procedural skills, knowledge of how science works, other)?
- How do you use your science assessment data?

Leadership

- Does your school use self-evaluation to inform strategic decisions in developing science provision? How?
- Do teachers in your school have access to people with expertise in science education? Who? How? How often?
- How often have staff you are responsible for accessed science specific CPD in the past 3 years? How did you identify appropriate CPD providers?
- What do you think are the sorts of challenges that teachers in your school experience with the teaching of science? Are there any specific to the type of school you work in/lead? How have you successfully overcome these?
- What do you believe best supports teachers' practice in primary science in your school?
- Are teachers in your school advantaged in any way in teaching primary science? Do these advantages come as a result of the specific type of school you work in/lead? How have these advantages come about and what do they result in?
- Is the Science lead in your school given dedicated leadership time to lead on this subject? How much/week?

Closing questions

- How have things been different as a result of the pandemic - are there any aspects of primary science practice from the past 18 months that you will retain, and any that you are keen to say goodbye to?
- Have you used the EEF Secondary Science guidance report to inform your practice? Which aspects, and why?
- How do you feel about leading primary science?
- What would be included in a good guidance report for primary science?
- Is there anything that we have not discussed that you think it is important that we include in our review?

Appendix 3: Draft extraction tool fields

The data extraction tool will include the following fields:

- Publication information: type, reference
- Geographical location of study: country
- Educational setting
- Sample characteristics: age, sex, indicators of socio-economic status, reported special educational needs
- Intervention/approach: name, description, objectives, who responsible, relationship with evaluator, group size, who delivered, duration, frequency, length of sessions, when delivered, reported outcomes
- Research design: assignment of participants, level of assignment, details of randomisation,
- Sample: number of school/class/participants in intervention/comparison groups, baseline characteristics, comparability between groups, attrition
- Outcome details: descriptive statistics reported, effect size
- Risk of bias arising from the randomization process
- Risk of bias due to deviations from the intended interventions: assignment to intervention, adherence to intervention
- Missing outcome data
- Risk of bias in measurement of the outcome
- Risk of bias in selection of the reported result
- Validity: how realistic

Appendix 4: Assessing the quality of quantitative studies

Assessing the quality of **quantitative** studies will focus on four areas: study design, features of data collection, features of analysis, and results and conclusions.

Study design

- Rationale for the study, justification of design, purposes of different methods or data sources, justification of time frame, study location
- Type of study
- Variables to be examined
- Key characteristics of the sample (sample size, age, gender, ability, socio-economic background, any other features of note)
- Methods used to identify sample, sampling frame, and sample selection
- Rationale for the selection of sample (basis for in/exclusion, discussion of sample size)
- Relationship of sample to population of interest
- Identification of limitations

Features of data collection

- Information about data collection methods
- Details of data collection tools
- Steps taken to increase reliability and validity of data collection instruments

Features of analysis

- Information about methods used to analyse the data
- Information about statistical methods used in the analysis
- Rationale for the methods of analysis for the study
- Information about reliability and validity of data analysis
- Information about strategies to control for bias from confounding variables

Results and conclusions

- Presentation of results
- Relationship of results to study research questions
- Identification of limitations

Appendix 5: Assessing the quality of qualitative studies

Assessing the quality of **qualitative** studies will focus on four areas: study design, features of data collection, features of analysis, and results and conclusions.

Study design

- Rationale for the study, justification of design, purposes of different methods or data sources, justification of time frame, study location
- Key characteristics of the sample (sample size, age, gender, ability, socio-economic background, any other features of note)
- Relationship of sample to population of interest
- Rationale for the selection of the target sample (basis for in/exclusion, discussion of sample size, selection of settings, etc.)

Features of data collection

- Information about data collection methods (e.g. for interviews/conversations: audio or video recordings; for field notes: conventions for recording events and/or distinguishing between description and comment/analysis)
- Information about data collection instruments (e.g. interview schedules, observation sheets, field notes routines, diary keeping instructions, etc.)
- Steps taken to increase reliability and validity of data collection instruments
- Information about measures for increasing trustworthiness of the data collected (e.g. influence of fieldwork methods/settings on the nature of the data collected; features of data indicating depth, detail, richness; etc.)

Features of analysis

- Information about how descriptive analytic categories, classes, typologies, etc. have been generated and used
- Information about the context of data sources?
- Information about diversity in the data? (e.g. multiple perspectives, alternate positions, negative cases, outliers, exceptions)
- Information about patterns of association or linkages with divergent positions or groups in the data, including possible explanations
- Information about measures for increasing trustworthiness of the analysis process (e.g. exploration of the detail, depth and complexity, concepts and meanings, discussion of explicit and implicit explanations, discussion of underlying factors/influences; discussion of patterns of association/conceptual linkages within data; presentation of illuminating extracts/observations)
- Information about any corroborating evidence used (e.g. other data sources or research evidence used to support or refine findings)
- Information about the generation of criteria for effectiveness or impact
- Information about how evaluative judgements have been reached

- Reflection on unintended consequences of intervention

Results and conclusions

- Presentation of results
- Relationship of results to study research questions
- Identification of limitations