Review of SES and Science Learning in Formal Educational Settings

A Report Prepared for the EEF and the Royal Society

September 2017

Terezinha Nunes, Peter Bryant, Steve Strand, Judith Hillier, Rossana Barros and Jaimie Miller-Friedmann

Contents

List of figures

List of boxes

List of tables

<div align="center">**Executive summary**</div>

The problem

Contemporary Western societies depend on there being widespread knowledge of science and understanding of scientific methods among their populations. This is because scientific skills are increasingly in demand in industries and in public services, such as in our health systems, and also because rational discussions about many current political issues that affect everyone, such as decisions about the environment, depend on people in general understanding the nature and the importance of scientific evidence.

Another reason for paying attention to the spread of scientific knowledge in the UK and in other countries is that scientific training eventually opens the door for young adults to many interesting and valued occupations. It is a matter of social justice that these opportunities should be widely available to everyone.

Unfortunately, existing research on pupils' attainment in science in U.K. schools has consistently shown an uneven spread of scientific knowledge. There is a consistent link between pupils' socio-economic status (SES) and their attainment and participation in science learning at school. Pupils who come from higher SES families are more likely to do well in science subjects than less advantaged pupils and to continue to study science after the age of 16 years, when it is no longer compulsory to do so.

Even among the pupils who do decide to continue with science at school after the age of 16, those from relatively high SES backgrounds still do a great deal better in science examinations than pupils from poorer homes.

A strong relation between pupils' SES and their attainments in science learning has also been found in many other countries in the world: the relation is apparent from the earliest age at which pupils' knowledge and understanding of science is assessed and it continues throughout pupils' years at school.

The extent of the problem in the UK

This report contains an extensive analysis of data in the UK National Pupil Database (NPD) on the performance of disadvantaged pupils in national science tests and in tests of other subjects in comparison to those of pupils from higher SES backgrounds. This analysis confirmed that pupils from economically disadvantaged families (pupils who have been entitled to Free School Meals at least once in the last six years) have much lower scores in national science tests and examinations (Key Stages 1, 2, 4 & 5, A level) than pupils from higher SES families.

The new analysis also showed that disadvantaged pupils make poor progress in science at every stage of their school career. Whenever we looked at the differences in science attainment between disadvantaged and other pupils at one stage after controlling for the pupils' level of attainment at previous stages, we found that there was still a gap between the two groups in their progress in science. Even when we took account of, and controlled for, their earlier difficulties, disadvantaged pupils still made less progress than other pupils. The gaps grow particularly strongly between ages 5-7 and ages 11-16, which coincide with particularly significant times in cognitive development.

The same gap between disadvantaged and other pupils shows up in figures for participation in science after it ceases to be a compulsory subject at school. The analysis of the NPD data amply

confirms previous research which had shown that disadvantaged pupils are proportionally less likely than other pupils to continue with science in the post-16 years.

The gap between disadvantaged and non-disadvantaged pupils is not unique to learning science. The NPD data show equivalent gaps between the two groups of pupils in their attainments in other subjects, such as English and mathematics, as well. The generality of the difficulties that disadvantaged pupils have in succeeding at school is an obvious matter for concern, but it is valuable information also for researchers who are trying to explain the strong connection between pupils' SES and their science learning. It means that the factors that hold low SES pupils back in school attainment are likely to be ones that affect a wide range of school outcomes.

The 2015 PISA (Performance Indicators on Student Assessment) data can be used to contextualise the size of the problem in the UK, because the measures of SES and science attainment are the same across countries. The amount of variance explained by SES in science scores in the 2015 PISA decreased in the UK since the last PISA that focused on science in 2006; it is currently 10.5%, which is less than the average figure for OECD (Organisation for Economic Co-operation and Development) countries, which was 12.9%. This figure is comparable to the percentage of variance explained by SES in Finland (10%) and lower than the figure for the United States (11.4%). At the same time, the proportion of students at the bottom SES level who perform comparably to those at the top SES level increased by 5%, to 35%.

What causes the SES-science attainment gap?

The link between SES and science attainment naturally prompts the question whether there are any identifiable factors in children's and adolescents' environments and experiences at home and at school that cause or exacerbate these differences. To put the question more technically: are there any possible variables that mediate the evident effects of SES differences in science learning? It is obviously important to learn about the variables that mediate the SES and science attainment effect for theoretical reasons, and also for practical, educational reasons. If we can discover what these variables are, we may also be able to make them part of an educational programme to improve pupils' science learning.

The research carried out for this report led to the identification of some possible explanations and to the elimination of other explanations that had been considered plausible in the past.

a. Opportunities to learn

One possible hypothesis is that low SES pupils are held back by a lack of opportunities for learning because of the restricted financial circumstances of their family life or the poor resources for teaching science in schools in areas of deprivation. There is evidence for an "opportunity gap" that is related to school SES level. The SES level of different schools, which is measured by taking the average SES of the pupils in each school, accounts for variance in children's science attainment, even after the relation between the individual pupils' SES and their science attainments is taken into account. Strong evidence for the school SES effect has been found in the PISA data in Australia and in the ALSPAC data (Avon Longitudinal Study of Parents and Children) in the UK. Pupils from lower SES backgrounds perform significantly better in science assessments if they attend schools with a higher SES level; conversely, students from higher SES backgrounds perform less well than their peers if they attend schools with lower SES (although this did not apply to students at the highest SES level in the ALSPAC data).

One plausible explanation for this link between the school SES level and pupil attainment in science is that resources for teaching science are also related to the school SES level. Factors such as how well science laboratories are equipped, teacher qualification and availability, and amount of time invested in science activities are school factors related to science attainment.

There is also an opportunity gap in families: pupils from lower SES backgrounds have less access than pupils from higher SES backgrounds to educational resources (e.g. desks, dictionaries); access to these resources, in turn, has been found to predict science attainment.

b. Interest in science

Several researchers have entertained the idea that the interest that pupils have in science affects how well they learn the subject and also whether they continue to take science courses right through school. In fact, there is very little in the way of empirical results to support this suggestion. Research within countries has at best shown only a very modest link between pupils' interest in science and their science attainments. In international comparisons, pupils in countries whose attainments in school science tests are relatively high actually show less interest in science on average than pupils in countries in which pupils do relatively poorly in the same tests.

c. Cognitive mediators

In our search for intervening variables that might account for the difficulties that low SES pupils have in science attainment, we also turned to factors whose possible link to SES is not as obvious as it is with the variables that we have discussed so far. These are *cognitive skills*, such as the child's ability to reason logically, that are known to play a significant part in pupils' science learning. It seemed possible that one or more of these cognitive variables might act as a mediator between pupils' SES levels and their attainments in science. For example, low SES pupils might fall behind in science because they are less able than high SES pupils to reason effectively. To identify such mediators would be extremely valuable educationally, because cognitive skills are often teachable, and any improvement in genuine mediators of the SES-science attainment link could well diminish or even demolish the SES gap in science learning.

To be accepted as a genuine mediator of link between SES and science learning, the variable has to satisfy three requirements:

1. The cognitive variable (e.g. reasoning) must be related both to pupils' SES levels and to their science attainment. The most convincing form of evidence on the relation between the cognitive variable and science attainment is longitudinal data which includes a measure of the possible mediator earlier on and a measure of science attainment later. The evidence will be even more convincing if the relation between the mediator and science attainment continues to be significant after the effects of a third variable, which could be a causal factor of both the mediator and science attainment - such as measured intelligence - is taken into account.

2. When the cognitive variable is entered into a regression analysis together with SES and a measure of science attainment, it should have the effect of reducing the strength of the relation between SES and the pupils' science attainment. In other words, if you take account of the cognitive variable and control for differences between the pupils in measures of this variable, the relation between their SES levels and their science attainment will be much weaker than if you do not.

3. Any intervention that improves pupils' performance in measures of the specific cognitive skill that is hypothesized to be a mediator should raise the level of their science attainments as well.

Ideally, a convincing test of the hypothesis that a particular cognitive ability acts as a mediator of the link between SES and science attainment should tackle all three requirements in one and the same study. After a lengthy and very thorough search of published research on the subject, we concluded that no such comprehensive study has been done. However, because we looked at the work of a wide range of researchers, we could identify three cognitive abilities which different researchers had shown to satisfy one of the three requirements, and their work combined appeared to satisfy all three requirements. This pattern of research is much less satisfactory than a single comprehensive study which looks at all three requirements. The report provides evidence relevant to the first two requirements in correlational studies by analysing data from ALSPAC: it includes analyses of specific cognitive skills that are longitudinal predictors of science attainment at a later date, investigates whether these cognitive skills reduce the amount of variance explained in science attainment by SES, and examines whether the link between the cognitive skills and SES could be explained by measured intelligence. Because ALSPAC is a longitudinal study, it does not address the third requirement of showing that an intervention that improves pupils' performance in specific cognitive skills raises their attainment in science too.

The report puts forward three cognitive variables as likely mediators of the relation between SES and science learning. They are:

1. *Scientific reasoning*: particularly the ability to understand how causal variables should be isolated and varied independently from each other in experiments.   The literature search showed correlational evidence to support a link between scientific reasoning, SES, and science learning. The analyses of the ALSPAC data showed that a measure of scientific reasoning - the Control of Variables Task - is a longitudinal predictor of science attainment and reduces the amount of variance explained by SES in science attainment. The analyses also showed that the mediating role of scientific reasoning between SES and science attainment is independent of measured intelligence.

2. *Literacy:* in correlational studies of science learning, the strongest and most consistent predictor of pupils' scientific attainment has undoubtedly been how li*terate* they are. Some of the possible reasons that have been given for this connection are the importance of reading scientific texts and preparing written scientific reports; the effects of reading on pupils' scientific vocabulary; the usefulness of understanding the morphemic structure of words in learning scientific terms. There is a strong relationship between pupils' SES and their literacy. The analyses of the ALSPAC data showed that a standardised measure of reading comprehension is a longitudinal predictor of science attainment and reduces the amount of variance explained by SES in science attainment. The analyses also showed that the mediating role of reading comprehension between SES and science attainment is independent of measured intelligence.

   A combination of scientific reasoning and reading comprehension as joint mediators of the SES-science attainment link reduced the predictive power of SES to negligible values: the amounts of variance explained by SES in science attainment varied between 0.8% and 2.1% in the different Key Stage assessments. This new evidence comes as

welcome news, since there are already tried and tested ways of improving both these cognitive skills.

3. *Metacognitive ability*: pupils' ability to think about their own cognitive activities is related to their success in science, as well as to their SES levels, and an intervention that involved metacognitive training did have an effect on pupils' science attainment. This kind of training has been shown to help lower attaining learners more than other pupils.

Educational programmes aimed at improving science learning among low SES pupils

The report ends with a review of recent educational programmes for low SES pupils. These tend to be multi-faceted in that the teaching in these programmes is usually aimed at a number of variables. The success of the different programmes varies, and appears to be strongly influenced by the quality of the professional development provision in each project.

Most of the successful educational programmes involve teaching one or more of the three possible mediators of the SES-science learning relation which we have already described. Other programmes have had a socio-cultural focus, which takes one of the following forms:

1. Bringing students into a science 'place' e.g. university laboratories or a science museum

2. Bringing scientists or extra-curricular science activities into schools

3. Developing teachers' understanding of students' perspectives

Programmes seeking to develop pupils' skills in scientific reasoning, literacy and metacognition tend to be given to pupils in upper primary school or lower secondary school and have had positive effects. Programmes with a socio-cultural emphasis, which aim to break down perceived barriers between pupils and science, tend to be given to pupils in secondary school, particularly older pupils, again with beneficial results.

Implications for future research and education

The review of the literature and analyses carried out for this report leave no doubt that the variables identified here as possible mediators of the link between SES and science should be investigated in studies that combine longitudinal and intervention research. Such studies are not difficult to carry out. A large scale randomised controlled trial can offer the opportunity for this combination because the comparison between the intervention and the control group provides evidence regarding whether improvements in the cognitive variable lead to improvements in science attainment while the analysis of pre- and post-test data in the control group can provide longitudinal evidence for the mediating role of the cognitive variable. In order for such intervention studies to be successful in collecting all the necessary data for the combination of longitudinal and intervention methods, pre- and post-tests must be carefully chosen. Failure to measure both the cognitive variable and science attainment at pre- and post-test, and to relate these measures to SES, would compromise the test of the mediator role of the cognitive variable. This combination of longitudinal and intervention data in the same study is rare, but it has been implemented to investigate causal connections both with literacy and mathematics learning.

It is often suggested that early intervention is the best solution. However, as the analysis of the NPD data identified two points in pupils' school career when the gap in science attainment increases - at about 5-7 and 11-16 - it is suggested that early intervention is not sufficient. These ages coincide with the timing of important changes in children's cognitive development, and it is unlikely that pre-school interventions would suffice for promoting the cognitive skills that are mastered later in children's lives. Interventions relevant to these time points should be prioritised in research.

The literature on science learning has several other hypotheses about classroom experiences that promote pupils' progress in science learning: for example, cognitive conflict in group or paired work, practical experimental work and argumentation in writing scientific reports. Some of these hypotheses have been tested with respect to specific scientific concepts (e.g. cognitive conflict and the trajectory of falling objects). It is of great interest to investigate whether such experiences promote specifically the understanding of the concept targeted in the lessons or whether they promote scientific reasoning and knowledge in a more general way.

The evidence about the three cognitive mediators analysed in this report has been gathered in a piece-meal manner, but there is enough of it now to make us believe that teachers and teaching programmes should incorporate into their science lessons activities that improve students' reasoning about the importance of controlling variables in scientific experiments as well as activities that improve their literacy and meta-cognitive skills. Research on the effects of taking these steps might very well show that they reduce the relation between SES and science attainment by promoting science learning by disadvantaged pupils

Chapter 1

Introduction[1]

> Box 1 Chapter 1 Summary
>
> The Education Endowment Foundation and the Royal Society commissioned a report to update and extend a previous report by the Royal Society and provide answers to three questions: (1) How can the link between socio-economic status (SES) and attainment and participation in science be described? (2) What are the possible causes of this link between SES and science attainment and participation? (3) What are promising pedagogies and programmes that are likely to impact on the attainment and progression of disadvantaged students in science subjects?
>
> The evidence of a link between pupils' socio-economic background and their attainment and participation in science is robust. Some of the reasons for addressing this link are the aim of achieving equity in education, the ever growing need for professionals with a background in science, and the significance of scientific literacy for all citizens in today's society.
>
> - This review addressed the questions in the terms of reference by conducting secondary data analyses on two large data sets - the National Pupils Database for England (NPD) and the Avon Longitudinal Study of Parents and Children (ALSPAC) - and by carrying out three systematic reviews of the literature.
> - The analyses of the NPD examine attainment and participation. The three literature reviews in this report focus only on attainment because there was modest scope for expanding the conclusions from the previous report with respect to participation.
> - With respect to participation, the main finding from the previous report, replicated in this review, is that previous attainment is the major predictor of later participation. The current review also found that:
>   1. whereas previous mathematics and science attainment are positively related to later participation in science, previous language attainment in school tends to be negatively associated with later participation in science (i.e. higher attainment in language is associated with less participation in science);
>   2. schools can have an effect in increasing participation by offering science courses before science is compulsory and making the relevance of science clear in lessons, among other measures;
>   3. policies that aim at widening participation have not proportionally increased participation by pupils from lower SES backgrounds in science courses.

Aims of the report

The Education Endowment Foundation and the Royal Society commissioned a review of the relation between socio-economic background (SES) and attainment and participation in science stipulating that the purpose of the review is to:

---

[1] This chapter was prepared by Terezinha Nunes, Peter Bryant and Rossana Barros.

1. identify the link between disadvantage and attainment and participation in science subjects;
2. explore the cause of any attainment or participation gap with the aim of helping the EEF select projects that are likely to address these causes (e.g. parental expectations or pupil disaffection); and
3. identify promising pedagogies, interventions and programmes, within school and/or involving families, that are likely to impact on the attainment and progression of disadvantaged students in science subjects.

Background

In the UK, as in many other countries, pupils from prosperous homes show higher attainment and greater participation in science courses when these become optional than pupils from less prosperous family backgrounds. Despite the recent expansion in higher education and multiple initiatives to change school science courses, participation in science courses by students from lower SES backgrounds is still relatively low. These highly consistent differences raise the question whether there are things that can be done to remove or at least to diminish this disadvantage; this is a question for researchers to answer.

Among the many reasons for aiming to reduce the disadvantage faced by pupils from less prosperous family backgrounds in attainment and participation in science are: (1) the aim of achieving equity in education, (2), the ever growing need for professionals with a science background, which might only be met if pupils from all SES backgrounds increase their participation in science learning, and (3) the significance of scientific literacy in today's world.

Scientific literacy is crucial for citizens today as everyone is, now more than ever before, involved in using scientific information to make decisions about their own future and that of the next generation. Some decisions may be more personal, such as maintaining a healthy diet or opting to take medication after considering its benefits and risks, whereas others have significance beyond the personal sphere, such as how far one contributes to the management of waste and to mitigating the effects of global warming. It is therefore imperative that scientific literacy becomes part of the taken-for-granted aims of education for all.

A previous report by the Royal Society (2008) showed a negative relation between SES and science attainment at Key Stages 1, 2 and 4. It also established the existence of a participation gap, as disadvantaged students are less likely to study separate science courses when these become optional and more likely to study combined, dual, single and general science courses, which do not prepare them as well for continuing to study science. The report also acknowledges that the participation gap may be largely explained by prior attainment. The negative association between SES and attainment does not appear to be larger in science than in other subjects, but actions specific to science education may be required if the negative association between SES and science attainment is to be weakened.

About a decade has passed since the Royal Society 2008 report. The Education Endowment Foundation (EEF) and the Royal Society commissioned this new report to update and extend the previous report where possible, stipulating that the new report would present a review of the relevant research and an analysis of existing data sets on UK pupils.

In the subsequent sections, this introduction presents the terms of reference by describing the aims of the report, the scope of the review of the literature included in the present report, and the ways in which this report sought to update and extend the 2008 report.

The scope of the review

In order to address these aims, we carried out secondary data analyses of two data sets and three systematic reviews of the literature. The data sets were provided by the National Pupil Database (NPD) for England, which contains data on attainment and participation, and by the Avon Longitudinal Study of Parents and Children (ALSPAC), which contains longitudinal data on science attainment as well as a rich source of information on pupils' SES background and their cognitive development.

All three reviews of the literature include national and international publications. Although one must be cautious about applying research conclusions from one country to another, the international literature can be used to clarify whether the SES-science learning relation is stronger or weaker in different contexts and how it is explained in other countries. Measures of SES as well as measures of science attainment vary across studies (for a discussion of measures of SES, see Appendix 1.1), but they generally address the same aims. The National Curriculum's (NC) aims for science education in the UK and the aims of measures of scientific literacy in the Programme for International Student Assessment (PISA), which were used in much of the literature reviewed in this report, are clearly linked conceptually (see Appendix 1.2 for NC aims for science and the definition of scientific literacy in PISA).

The present report focuses on attainment as well as participation in the analysis of the NPD. However, the report does not include a detailed review of the literature on participation because the same flaws in studies of participation noted by Gorard and See (2009) appear in studies published after their review. A detailed analysis of the literature on participation is not included here because there was little scope for extending the findings of the 2008 review. The main new findings are presented briefly, but we first reiterate the conclusion of the 2008 review: previous attainment is a robust predictor of participation in science and the main obstacle to participation by pupils from lower SES backgrounds. Once attainment is taken into account, the effect of SES on participation is negligible quantitatively. Therefore, the key to increasing participation is to improve pupils' attainment before they have to opt for taking science courses.

Attainment in mathematics and in science during compulsory education are **positive** predictors of later participation in science; in contrast, at least two studies (one of them in England; Homer, Ryder, & Banner, 2014) have found that attainment in language is a **negative** predictor of participation in science, a result that was not brought out by the previous review. Further replications are desirable as the effect was unstable in one study.

The majority of pupils eligible for free school meals who attain a good level of performance in science at the end of compulsory education continues to study science in the subsequent two years; these pupils are over-represented in the Applied Sciences and in the Business and Technology Education Council awards and under-represented in other awards. Homer, Ryder, and Banner (2014) conjecture that science awards that are more likely to translate into employment early on might be preferred by pupils eligible for free school meals.

Schools have an effect on participation in science, which is independent of the SES effect. Characteristics of schools that have higher levels of participation in science are the offer of science subjects before they are compulsory, time tabling that facilitates taking science courses, the offer of science and vocational subjects that may work well together, teaching science in ways that make it seem interesting and useful, and including real-life applications in science lessons. It must be noted that these analyses of associations between school

characteristics and participation are correlational; further research that uses intervention methods are required for robust conclusions (for a summary of these studies, see Appendix 1.3).

New policies have been introduced in England (and elsewhere) to widen participation, but these did not proportionally increase the participation of students from lower SES backgrounds in science, in spite of their positive effect on participation in higher education in general (see Appendix 1.4 for a brief review of policies).

The structure and contributions of the present report to understanding the link between SES and science attainment

The way in which this report addresses the aims set out in the terms of reference and how it updated and extended the 2008 Royal Society report are briefly described here.

1. Identifying the link between disadvantage and attainment and participation in science subjects.

In order to address this aim, two data sets were analysed in this report: data from the National Pupil Database for England (NPD) and data from the Avon Longitudinal Study of Parents and Children (ALSPAC).

The analysis of the NPD, presented in Chapter 2, updates the previous review as it uses the 2015 attainment scores and extends the previous report by inspecting the associations between SES and pupils' progress during compulsory education. Progress is defined in the analyses by considering whether the scores obtained in later assessments (e.g. Key Stage 1) are in line with those expected from attainment in earlier assessments (e.g. assessment at the end of Reception). If the attainment in later assessments is in line with what was expected on the basis of earlier assessments, then the SES gap does not close, but progress in science and SES are not related. In contrast, if attainment in later measures is below what was expected from the earlier measures, then SES is related to the initial gap in attainment as well as to progress in science learning. Finally, if the attainment in later assessments is above what was expected, the SES gap closes and SES is negatively related to progress in science attainment (i.e. pupils from lower SES make more progress than those from higher SES).

In relation to post-compulsory education, the analyses in the current report take into account that there are socio-economic variations in participation in education post-16 as well as variations in A level science achievement. The analyses in the current report identify how far A level science achievement gaps are driven by these varied patterns of participation in education post-16 and whether they continue to develop during the age 16-19 phase.

The analyses of the ALSPAC data, presented in Chapter 5, extend the findings from the previous report in two ways. First, a different measure of SES is used, the highest level of education attained by the mother. In contrast to eligibility for free school meals, which is a binary measure (yes or no), mothers' education provides an ordinal scale and is considered in the literature (Gottfried, Gottfried, Bauthurst, Guerin, & Parramore, 2012) a more robust index (for an overview of measures of SES in education, see Appendix 1.1). Second, the analyses implemented here calculated the school SES and investigated whether, in the UK as in other

countries, school SES and individual SES make independent contributions to the prediction of science attainment.

2. To explore the cause of any attainment gap with the aim of helping the EEF select projects that are likely to address these causes.

This report updates and extends the 2008 report by approaching the investigation of causes of the SES gap in science learning (i.e. achievement and participation) in three ways.

The first approach was to review research that includes data both on SES and on science learning, as well as an explicit hypothesis about what might cause the association between SES and science learning. The correlation between SES and science learning prompts the question: how does parents' [2]socio-economic status translate into influences on their children's science learning? It is extremely likely that the reason for the correlation between SES and science attainment is that parents' SES relates to some other variable, which itself partly determines how well pupils do in science. This other variable is usually described as an intervening variable (see Appendix 1.5 for definitions of different types of intervening variables, mediators and moderators), which is considered the cause of the association between SES and science learning. The research that contained an explicit hypothesis about what mediates the association between SES and science learning suggested two plausible hypotheses, which were termed in this report "the opportunity gap" and "the interest gap". The review of this research is presented in Chapter 3.

The second approach to exploring the causes of the attainment gap extends the previous review by using a new, but still rigorous, method of reviewing the literature. Here the review brings together different pieces of research that either provide data on how children learn science, irrespective of their SES background, or provide data on SES effects on other cognitive measures that are themselves related to children's science learning. For example, pupils' SES is related to their attainment in literacy and it is a plausible hypothesis that literacy could be an intervening variable that explains why SES and science attainment are related. Thus we sought to identify separate studies that would allow us to examine this hypothesis. Some studies would provide data on whether science learning is related to literacy and others would provide data on whether there is an SES gap in the specific literacy skills related to science learning. We recognise that the causal role of the cognitive factors identified in this way still remains a matter of conjecture, but these different sources of evidence would make their mediating role a plausible one. The review of these studies is presented in Chapter 4.

In our third approach to the question of causes of the SES-science learning gap, we turned to the analysis of the rich dataset provided by ALSPAC, which contains measures of some factors identified in Chapter 4 as plausible mediators of the SES-science learning achievement as well as measures of SES and of science achievement. Our approach was to examine whether, when these plausible mediators of the SES-science attainment gap were taken into account, SES still continued to be significantly related to science attainment. The results of this analysis extend the 2008 Royal Society report, which did not include any measures of possible causes of the SES-related attainment gap. The results of these analyses are presented in Chapter 5.

---

[2] Although papers refer to parents, it is usually the case that some measures are applied to caregivers, i.e. the person with whom the child lives if the child does not live with the parents.

3. To identify promising pedagogies, interventions and programmes, within school and/or involving families, that are likely to impact on the attainment and progression of disadvantaged students in science subjects.

The present review extends the 2008 report significantly by examining research that aimed to evaluate pedagogies, interventions and programmes that impact positively on disadvantaged students' learning. Different types of intervention were examined, including interventions that focus on specific topics (e.g. electricity, buoyancy) and others that focus on developing students' thinking skills in science (e.g. 'Let's Think', which builds on CASE, 'Cognitive Acceleration through Science Education'). This work was not reviewed in the previous report, which acknowledged the need to examine the extent to which specified teaching methods can result in better attainment in science by pupils from lower SES backgrounds.

Chapter 2

Identifying the link between socio-economic disadvantage and participation and attainment in science: An analysis of the England National Pupil Database (NPD)[3]

Box 2 Chapter 2 Summary

Chapter 2 presents up-to-date information about the relation between pupils' socio-economic status (SES) and their participation and achievement in science. We use EVER6 (whether the student has been entitled to a Free School Meal (FSM) at any time in the last six years) as the main measure of socio-economic disadvantage.

- The EVER6 achievement gaps in Science are large. The odds of achieving conventional benchmarks of success in science are much greater for NonFSM students than for EVER6 students;
- The achievement gap in science grows over time. The gap seems to be largest at the end of secondary school (age 16).
- The substantial EVER6 gaps in science at age 18/19 are largely driven by low achievement in science at age 16 and low levels of participation in Full-time Education (FTE) post 16.
- Among students who continue in FTE post-16, those who were EVER6 at age 16 still make less progress and have a lower average 'A' level science score than their NonFSM peers.
- The EVER6 achievement gaps in science are broadly similar in size to the achievement gaps in other subjects such as English and mathematics and in overall score. Science is not a special case.
- The earliest measure of achievement in science, recorded at age 7, is strongly predictive of later science achievement, but is itself strongly predicted by pupil's Early Years Foundation Stage Profile (EYFSP) score at age 5.

Introduction

This element of the review consists of secondary data analysis of the England National Pupil Database (NPD). It updates and extends the analyses in the Royal Society (2008) report 'SES and Science Education'. It analyses national data on science participation and achievement in England at Key Stage 1 (age 7), Key Stage 2 (age 11), in GCSE examinations (age 16) and through to 'A' level and other level 3 examinations (age 18/19). The key indicator of Socio-economic Status employed is whether a pupil has been entitled to a Free School Meal (FSM) at any time in the last six years (the EVER6 measure).

The purpose of the analyses are to:

- Compare the size of the EVER6 achievement gap over the school period from Foundation Stage to the end of Key Stage 5 using the most recent 2015 cross-sectional data;
- Evaluate EVER6 gap in progress (both in science and overall achievement) within each Key Stage;
- Where available chart trends in the size of the EVER6 gap over time;

---

[3] This chapter was prepared by Steve Strand.

- Compare the size of the EVER6 achievement gaps in Science to the size of the gap in other subjects;
- Breakdown the EVER6 gaps in Science at KS4 by other student and school characteristics, e.g. by gender, ethnicity, school type and region.

The measure of Socio-economic Status (SES)

There is substantial and long-standing evidence of a medium to strong association between socio-economic status (SES) and educational achievement, see for example the literature review of White (1982) and the meta-analysis of Sirin (2005). For example Sirin (2005) reports a moderate effect size (Pearson's r) of around 0.31, although the effect varies by the type of SES measure, the type of achievement measure and a range of student variables such as age and ethnicity.

There are different indicators of socio-economic status (SES) employed in the academic literature. These focus on different dimensions such as parental occupation, household income, parents' highest educational qualifications or family social resources. Many of these indices require detailed interviews with parents to ascertain reliable measures and are not available in the school context. Other measures of SES are area based such as the England Income Deprivation Affecting Children Index (IDACI) which measures the proportion of children under 16 in a neighbourhood[4] who live in households entitled to state benefits. However these have the problem that they are area based and do not necessarily reflect the circumstances in the family in which the pupil resides.

The main indicator which has been used in school based research is whether the pupils is entitled to a Free School Meal (FSM). Entitlement to a FSM is based on low income households defined as:

- Children in Income Support households

- Children in Income Based job Seekers Allowance households

- Children in Working Families Tax Credit households whose income (excluding housing benefits) is below 60% of median before housing costs

- Children in Disabled Person's Tax Credit households whose income (excluding housing benefits) is below 60% of median before housing costs

- National Asylum Support Service (NASS) supported asylum seekers

In January 2012 the Department for Education (DFE) in England introduced the EVER6 measure, which is an indicator of whether a pupil has been entitled to a FSM at any time in the last six years. This increased the size of the group identified from around 18% to around 28% of the school age 5-16 population.

No single measure of any construct is ever perfect, and FSM and EVER6 both suffer from being binary measures, and therefore (i) prone to threshold effects around the cut-off for entitlement, and (ii) lacking in differentiation, particularly within the larger group not entitled to FSM group (Strand, 2014). However since EVER6 is widely used, is the indicator that directly

---

[4]. A neighbourhood for the IDACI is the lower super output area of which there are 32,482 in England in 2010, intended to be roughly equal in size each containing on average contains around 1,500 people.

attracts funding through the Pupil Premium (in 2015/16 £935 for every secondary school pupil and £1,320 for every primary school pupil in the EVER6 group), is the measure of deprivation used for reporting in the primary and secondary school performance tables, and is utilised by the EEF in its evaluations, this is the measure of SES employed in the current analysis of the NPD.

In the January 2013 school census 27.6% of pupils in Reception through to Y11 were identified through the EVER6 measure. This is larger than the proportion eligible at any particular point in time, which in the January 2013 school census was 18.2% across Reception-Y11. A detailed breakdown of the proportion of pupils identified as FSM and EVER6 by year group for 2013 is included in Appendix 2.1.

Achievement gaps and effect size measures

In this chapter we will report the size of achievement gaps between EVER6 and those who have never been entitled to a FSM during the last 6 years (NonFSM). We will do this in the units in which the data have been collected, which may be National Curriculum levels, test marks, GCSE grades, points scores, or the proportion achieving a particular threshold e.g. the percentage achieving Level 4 or above at age 11 or the percentage passing the EBacc in science at age 16. With so many different measures, and indeed changes over time where the same measure is used, it is therefore useful to have standardised measures of effect size. We will report Cohen's D for continuous measures and the Odds Ratio (OR) for binary measures. These effect size measures are described in detail in Appendix 2.2.

Achievement at Key Stage 1 (Age 7)

Table 2.1 presents results for pupils at the end of Key Stage 1 in the areas of reading, writing, mathematics and science, as well as KS1 average points score (APS). KS1 Science is assessed in whole National Curriculum (NC) levels (W, 1, 2, 3 or 4) making the measure relatively undifferentiated as there are only five possible values. Therefore we created a further KS1 Science measure by calculating each pupil's average level across the four science attainment targets (AT): Scientific enquiry; Life processes and living things; Materials and their properties; and Physical processes. Results are reported both as a mean National Curriculum (NC) level and as the proportion of pupils achieving Level 2 and above (L2+) and Level 3 and above (3+).

The main findings are:

- At the end of KS1 in 2015 over 160,000, or 25.6% of all pupils, were recorded as EVER6.

- The EVER6 achievement gaps are large. In Science the mean NC level for NonFSM pupils was 2.18 while for EVER6 pupils it was 1.94, giving a Cohen's D of 0.44. i.e. on average NonFSM pupils were scoring almost half a standard deviation higher than EVER6 pupils (or equally EVER6 pupils were scoring almost half a standard deviation below their nonFSM peers). Similarly 27% of NonFSM pupils achieved Level 3+ compared to just 12.2% of EVER6 pupils, indicating the odds of achieving Level 3+ were 2.7 times higher for NonFSM pupils compared to EVER6 pupils[5].

---

[5]. For a detailed explanation of how the effect sizes (Cohen's D and the Odds Ratio) are calculated see Appendix 2.

- The size of the EVER6 achievement gap is broadly consistent across reading, writing, mathematics and science at around 0.45 SD and with the odds of NonFSM pupils achieving Level 2+ (Level 2 and above) and Level 3+ (Level 3 and above) about 2.5 times higher than for EVER6 pupils. For average KS1 points score NonFSM pupils scored on average the overall half a standard deviation above EVER6 pupils (Cohen's D=0.50) and the odds of achieving Level 2+ or Level 3+ in all of reading, writing, mathematics and science were 2.3 times higher for NonFSM pupils than for EVER6 pupils.

- The EVER6 gap is already large at the end of the Reception year, pupils' first year in primary school. The Early Years Foundation Stage Profile (EYFSP) records teacher ratings across 17 areas of learning. EVER6 pupils on average score half a standard deviation lower than NonFSM Pupils (Cohen's D=0.50), and only 37% achieve a Good Level of Development (GLD) compared to 57% of nonFSM pupils, indicating the odds of achieving a GLD are 2.3 times higher for NonFSM than for EVER6 pupils.

- We take account of prior attainment through linear regressions for KS1 APS and KS1 Science score that control for pupil's EYFSP points score two years earlier. Controlling for prior attainment at the end of Reception Year reduces the EVER6 gaps by about half, but does not eliminate them. Thus EVER6 pupils are making less progress during the key stage than their NonFSM peers. In Science EVER6 pupils on average score -0.21 SD lower than predicted by their EYFSP score. However there is no evidence that this decline is specific to Science, since the EVER6 gap for progress in KS1 average points score is about the same size (D=0.26).

Table 2. 1 Achievement and progress at the end of Key Stage 1 by subject and EVER6 status: 2015

| | Never entitled to a FSM | | | | | Entitled FSM in last 6 years | | | | | Effect Sizes[a] | | |
| | Mean | | | % at level: | | Mean | | | % at level: | | Cohen's | OR | OR |
Assessment	score	N	SD	2+	3+	score	N	SD	2+	3+	D	2+	3+
KS1 Reading[b]	**17.1**	475,982	3.75	93.1	36.7	**15.2**	163,409	4.19	83.8	18.7	0.48	2.6	2.5
KS1 Writing[b]	**15.8**	475,983	3.67	90.8	20.6	**14.0**	163,409	3.93	79.1	8.7	0.48	2.6	2.7
KS1 Mathematics[b]	**16.9**	476,004	3.46	94.9	30.0	**15.2**	163,409	3.75	87.6	14.8	0.47	2.6	2.5
KS1 Science[b]	**2.2**	475,935	0.58	93.6	27.2	**1.9**	163,316	0.59	84.7	12.2	0.43	2.6	2.7
KS1 Science (AT aggregate)[c]	**2.18**	475,963	0.54	-	-	**1.9**	163,335	0.56	-	-	0.44	-	-
KS1 Avg. Points Score (APS)[d]	**16.5**	476,021	3.24	88.4	13.4	**14.8**	163,428	3.53	74.9	4.8	0.51	2.6	3.1
EYFSP points score[e]	**33.8**	460,027	7.22	57.0	54.1	**30.1**	158,666	7.35	37.0	34.1	0.50	2.3	2.3
KS1 APS - value added	**0.17**	459,869	2.36	-	-	**-0.47**	158,526	2.62	-	-	0.26	-	-
KS1 Science - value added	**0.02**	459,825	0.44	-	-	**-0.07**	158,457	0.46	-	-	0.21	-	-

Notes

(a) Cohens' D calculated using the SD for the entire population. OR= Odds Ratios (NonFSM : EVER6).

(b) Levels range from 0 (working towards level 1) through to level 4. For reading, writing and mathematics teachers are asked to make sub-levels judgments as to whether performance is just into Level 2, securely at level 2 or at the top end of level 2 (2C, 2B or 2A respectively). For science performance is scored as whole levels only.

(c) This measure is the average of the pupil's achievement across the four science attainment targets (Scientific enquiry; Life processes and living things; Materials and their properties; Physical processes). We use this measure as the science outcome in the Value Added calculation.

(d) The percentage achieving Level 2+ and Level 3+ reported are for success in all of reading, writing, mathematics and science.

(e) The Early Years Foundation Stage Profile (EYFSP) points score is the average across all 17 areas of learning. The two threshold measures reported are the % achieving a Good Level of Development and the % achieving at least the expected level in all 17 areas of learning.

(f) the correlation of EYFSP score with KS1 science was 0.57 and with KS1 APS was 0.68. The correlation between KS1 Science and KS1 APS was 0.88.

Achievement at Key Stage 2 (Age 11)

Table 2.2 presents results for pupils at the end of Key Stage 2 in the subject areas of English, mathematics and science, in the KS2 tests in reading, mathematics and grammar, punctuation & spelling (GPS) and for value added measures of progress age 7-11. KS2 Science is again assessed in whole National Curriculum (NC) levels, though these are now spread across a wider potential range from W (Working towards level 1) through to level 6. Individual Attainment Target data is not collected at KS2 and there are no KS2 Science tests, so the NC level teacher assessment (TA) is the only data available. For the subjects, results are reported both as a mean NC level and as the proportion achieving Level 4 and above and Level 5 and above. For the tests, the mean fine grade levels are reported along with a single summary measure of the proportion of pupils achieving a 'good' Level 4 (a Level 4B or above, definable only through test results).

The main findings are:

- At the end of KS2 in 2015 when they are aged 11 over 178,000 or 31.2% of all pupils were recorded as EVER6.

- The EVER6 achievement gaps are large. For example in Science the mean level for NonFSM pupils was 4.36 while for EVER6 pupils it was 4.00, giving a Cohen's D of 0.46. i.e. on average NonFSM pupils were scoring almost half a standard deviation higher than EVER6 pupils (or equally EVER6 pupils were scoring almost half a standard deviation below their nonFSM peers). Similarly 46.5% of NonFSM pupils achieved Level 5+ compared to just 24.9% of EVER6 pupils, indicating the odds of achieving level 5+ or above were 2.6 times greater for NonFSM pupils compared to EVER6 pupils.

- The size of the EVER6 achievement gap is broadly consistent across science, English and mathematics at around 0.46 SD and with the odds of NonFSM pupils achieving L4+ and L5+ about 2.5 times higher than for EVER6 pupils. Broadly comparable achievement gaps were also demonstrated for the test scores in reading, mathematics and grammar, punctuation and spelling (GPS). For KS2 average points score, Cohen's D= 0.49 and the odds of achieving Level 4B+ in all of reading, writing and mathematics are 2.3 times higher for NonFSM pupils than for EVER6 pupils.

- We have seen that the EVER6 gap was already large at age 7. We have taken account of prior attainment through linear regressions for KS2 average points score (APS) and KS2 Science TA to control for KS1 attainment four years earlier. Controlling for prior attainment at age 7 reduces the EVER6 gaps by about 80%, but does not eliminate them. Thus EVER6 pupils make less progress during KS2 than their NonFSM peers.

- The size of the EVER6 gap in progress, as well as achievement, is the same for average points score as it is for science, so the patterns are not specific to science.

Table 2. 2 Achievement and progress during Key Stage 2 by subject and EVER6 status: 2015

	Never entitled to a FSM					Entitled FSM in last 6 years					Effect Size		
	Mean			% at level:		Mean			% at level:		Cohen's	OR	OR
	Level	N	SD	4+	5+	Level	N	SD	4+	5+	D [a]	4+	5+
KS2 Teacher Assessment[b]													
KS2 Science TA	**4.36**	393,134	0.75	92.2	46.5	**4.00**	178,026	0.85	81.4	24.9	0.46	2.7	2.6
KS2 English TA	**4.41**	393,154	0.79	92.1	50.1	**4.03**	178,031	0.88	82.0	27.8	0.46	2.6	2.6
KS2 Maths TA	**4.52**	393,166	0.89	91.8	52.3	**4.09**	178,024	0.93	82.0	30.6	0.47	2.5	2.5
KS2 Tests[c]				%level 4B+					%level 4B+			% level 4B+	
KS2 Reading test (fine-grade)	**29.1**	393,213	4.5	84.8		**27.0**	178,087	5.4	70.6		0.44	2.3	
KS2 Maths test (fine-grade)	**29.7**	393,238	5.4	81.5		**27.1**	178,075	5.6	66.5		0.46	2.2	
KS2 GPS test (fine-grade)	**4.95**	392,717	0.9	77.9		**4.54**	177,434	1.0	62.0		0.45	2.2	
KS2 Avg. Points Score	**29.3**	393,130	4.6	74.9		**27.0**	177,990	5.1	56.1		0.49	2.3	
KS2 Value Added Scores													
KS1 Avg. Points Score	**16.0**	374,688	3.50			**13.8**	171,432	3.86			0.58		
KS2 APS - value added	**0.08**	374,502	2.73			**-.18**	171,205	3.17			0.09		
KS2 Science - value added	**0.02**	374,488	0.52			**-.04**	171,236	0.59			0.10		

Notes
(a) Cohens' D calculated using the SD for the entire population.
(b) Teacher Assessment (TA) levels are recorded in whole numbers ranging from 0 (working towards level 1) through to level 6.
(c) For the tests, marks are converted to fine grade levels. A single threshold of achieving a 'good' Level 4 (Level 4B) is used rather than separate Level 4 or above (4+)and Level 5 or above (5+)
thresholds. For the KS2 average points score(APS) the threshold measure shown is the percentage achieving level 4B or above (4B+) in all of the reading test, writing TA and mathematics test.

Achievement at Key Stage 4 (age 16)

The interpretation of performance in science at KS4 becomes more complex than at KS1 or KS2, since rather than a single national assessment completed by all students, different learning pathways can be followed and different qualifications taken. The DFE argue that it is compulsory for all state-funded schools to teach Science at KS4 (DFE, 2015, p8), though strictly this is true only in Local Authority (LA) maintained schools, since Free Schools and Academies are only required to "have regard" to the requirement[6]. A balance between the depth and breadth of science study has to be considered: how many discrete examinations should be taken, of what type and what proportion of curriculum time should be appropriated? As a result we consider participation and achievement in science across a number of levels:

- Whether students participate by taking at least one GCSE in a science subject, and their level of achievement at this threshold. Also what proportion of students take an alternative Level 2 science qualification such as BTEC First Certificate in Applied Science?

- Whether students take the combination of subjects that meet the requirements for the EBacc in Science. Specifically in 2015 whether a student is entered for GCSE examinations in core and additional science, in double science, or in three individual sciences (three out of biology, physics, chemistry or computer science), and whether they achieve two or more passes (A*-G) in these subjects.

- Whether students achieve the EBacc in Science, i.e. two or more GCSE Sciences at grades A*-C, and to give a differentiated measure their EBacc Science points score (their GCSE points score in their best two science subjects).

We also consider Progress in Science between age 11 and age 16, using the EBacc points score as the outcome and KS2 mean test score as the input.

The results of our analyses are presented in Table 2.3. The main findings are:

- We consider first whether students attempted any Science GCSEs and their achievement at the lower hurdle of 1+ GCSE in a science subject. 90.4% of NonFSM pupils entered 1+ GCSE science compared to 77.5% of EVER6 pupils. The odds of a NonFSM pupil entering a science GCSE where therefore 2.7 greater than for EVER6 students. NonFSM students were more likely to achieve a pass in a GCSE science (OR=2.9) and more likely to achieve a higher grade (A*-C) pass in a science subject (OR=3.2). If we consider the highest points score achieved in any Science qualification the difference was around 6 points or one GCSE Grade, a Cohen's D= 0.61.

- The above figures do not include BTEC qualifications, and EVER6 students (11.7%) were more than twice as likely as NonFSM students (6.5%) to study for a BTEC in applied science. The majority of those achieving a BTEC in Science took no other Science GCSEs (80%) so the omission of BTEC in the science GCSE measure has a proportionately greater influence on the Science achievement of EVER6 students. Nevertheless BTECs are not allowed to contribute to the EBACC science.

---

[6]. https://www.gov.uk/government/publications/national-curriculum-in-england-framework-for-key-stages-1-to-4/the-national-curriculum-in-england-framework-for-key-stages-1-to-4

Table 2. 3 Achievement and progress during Key Stage 4 by subject and EVER6: 2015

KS4 Science Indicator	Never Entitled FSM			Entitled FSM last 6 years			Effect Size (OR/D)
	Mean	ValidN	SD	Mean	ValidN	SD	
Attempted Science GCSE	90.4%	407,808	0.29	77.5%	152,422	0.42	2.7
Achieved A*-G in Science GCSE	90.0%	407,808	0.30	75.6%	152,422	0.43	2.9
Achieved A*-C in Science GCSE	69.5%	407,808	0.46	41.8%	152,422	0.49	3.2
Achieved equiv. of Level 2 in BTECs in Science	6.5%	407,808	0.25	11.7%	152,422	0.32	0.53
Highest point score in any Science subject (GCSE or equivalents)	42.8	369,919	9.26	36.7	118,704	10.6	0.61
Entered for three individual sciences	25.7%	407,808	0.44	10.2%	152,422	0.30	3.1
Entered EBacc (minimum two sciences)	79.8%	407,808	0.40	56.8%	152,422	0.50	3.0
Achieved two Sciences A*-G	79.5%	407,808	0.40	56.0%	152,422	0.50	3.1
Achieved EBacc (core & additional or double pathway)	34.3%	407,808	0.47	21.7%	152,422	0.41	1.9
Achieved EBacc (three separate sciences pathway)	24.1%	407,808	0.43	8.8%	152,422	0.28	3.3
Achieved EBacc (total all pathways)	58.4%	407,808	0.49	30.5%	152,422	0.46	3.2
EBacc Science points score (0 if none)	34.3	407,808	18.8	21.8	152,422	20.2	0.63
EBacc Science points score (entered only)	43.0	325,297	8.3	38.4	86,583	9.0	0.53
Best 8 points score (GCSE & equiv.)	329.1	408,136	89.4	256.9	152,716	113.7	0.71
Included in Science VA calculation	76.4%	311,775	0.42	54.7%	83,457	0.50	2.7
EBacc Science VA score	0.44	311,775	6.18	-1.64	83,457	7.21	0.32
Best 8 VA Score	16.62	311,787	52.8	2.06	83,460	65.8	0.26

Notes

- *All analyses based on Maintained Schools Only (values of 8, 10 and 11 for KS4_TOE_CODE excluded).*
- *EVER6: In the KS4 file 61,816 cases were missing for EVER6, the vast majority (almost 91%) being pupils from establishments were the census is not completed: Independent Schools (48,734), FE Colleges (5,691), Secure Units (151) and those with missing KS4_TOE_CODE (2,041). Excluding these institutions 5,199 cases were missing EVER6 or just 0.8% of the population.*
- *Effect Size: Shaded cells indicate Cohen's D, unshaded cells indicate Odds Ratio (OR).*
- *EBacc Science points score: This is the average points score of the students best two eligible GCSE sciences, with 0 entered for a missing score if there is one and 0 for those who were not entered for EBacc at all. Two scores are calculated, one for all KS4 students and one only for those entered for at least one EBacc science subject. An AS levels score will always be taken over a GCSE in the same subject but the point will be capped at 58 (equivalent of a GCSE A* grade). See this link in the RAISE Online library: https://www.raiseonline.org/OpenDocument.aspx?document=217*
- *EBacc Science VA score. This is calculated only for students who fulfilled the EBacc entry requirement. It regresses the EBacc score on the students KS2 prior attainment (average of the KS2 English and mathematics scores using fine-points).*
- *Best 8 VA score: for comparability with the EBacc Science VA score this is also calculated only for students who entered at least one EBacc science.*

- Close to four in every five (79.8%) of NonFSM pupils are entered for EBacc science compared to just over half (56.8%) of EVER6 students, making the odds of NonFSM pupils being entered for EBacc science three times higher than the odds for EVER6 students (OR=3.0). The under-representation of EVER6 students is not just about entering three separate science subjects, it is about the lower proportion entering at least two sciences primarily through the core and additional science route.

- To achieve the EBacc Science a student requires two GCSE science passes at grades A*-C. This is achieved by 58.4% of NonFSM but only by 30.5% of EVER6 students, making the odds of achieving the EBacc science 3.2 times higher for NonFSM students than EVER6 students. NonFSM pupils were also more likely to achieve their EBacc in science through the separate sciences (25.7% entered and 24.1% achieved it through this route, compared to 10.2% entered and 8.8% achieved among EVER6 students).

- To reflect the average level achieved among all students an EBacc points score is calculated based on their two highest science subjects, with zero entered for a missing entry. The average score difference between EVER6 and NonFSM is 12 points, or just over two GCSE grades in each of the two science subjects, a Cohen's D of 0.63 and a large gap. This is a broadly similar size to the Best 8 point score gap (D=0.71) so is not specific to science. If the comparison is restricted just to those entered for an EBacc science subject the Cohen's D is slightly lower at D= 0.53.

- We calculated value added measures for EBACC science and for overall performance (Best 8 points score) through linear regressions controlling for KS2 average points score at age 11. To ensure we are comparing like with like, we compute the EBacc Science and the Best 8 value added measures just on those students who were entered for EBacc science. After controlling for prior attainment the EVER6 EBacc science gap reduces from D= 0.53 to D= 0.32, or by about 40%, reflecting the lower age 11 prior attainment of EVER6 students. This is still a large gap, indicating poorer progress even just among the 57% of EVER6 students who are entered for EBacc science. The poor progress is not specific to science since the achievement gap for Best 8 points score for the same students is D= 0.26, a comparable size.

- Overall we conclude that EVER6 gaps exist at all levels of science participation and achievement at KS4. EVER6 students are twice as likely to be entered for BTEC Science which is not included in the EBacc Science measure; they are 3.0 times less likely to be entered for the EBacc and 3.1 times less likely to achieve it; they are 3.3 times less likely to be entered for three individual sciences; their average EBacc science points score is 0.63 SD lower than their nonFSM peers. While around 40% of the EBacc science points gap can be explained by prior attainment 60% cannot, and reflects poorer progress in Science even among those EVER6 students who are entered for the EBacc. However the patterns observed for Science is similar to that observed for overall achievement so not specific to science.

KS4 Time Series 2013-2015

Appendix 2.3 includes the equivalents of Table 2.3 for 2013 and for 2014. We need to be somewhat circumspect in interpreting absolute figures in terms of achievement across time, particularly in recent years where major changes to the exam system have been enacted. For

example in 2014 two major reforms were introduced. First the Wolf Review restricted substantially the number of vocational qualifications that could be included in performance measures, prevented any vocational qualification counting as more than one GCSE, and capped the number of Non-GCSEs included in performance measures to two per pupil. Second, a new early entry policy meant that only a pupil's first attempt at as qualification rule counted in performance tables. This applied only to EBacc subjects in 2014 but has been extended to all subjects in 2015. In addition in 2015 Computer Science was added as one of the EBacc sciences for the first time.

Because of the above factors results are not necessarily directly comparable over time. However to the extent that we are interested in the gap between disadvantaged (EVER6) and non-disadvantaged students, we can still ask how this gap has changed over time, regardless of variation in the absolute level of achievement. Table 2.4 below presents the data on entry to and achievement in the EBacc in Science since its introduction in 2010 and includes the gap data we have calculated here for the last three years.

Table 2. 4 EBacc science entry and achievement 2010-2015

EBacc Science	2010	2011	2012	2013	2014	2015
% entered three sciences	16.5	20.1	23.2	24.7	22.2	21.7
% entered core & additional science[a]	46.8	41.4	40.7	41.5	46.5	52.7
% entered EBacc in total	63.3	61.5	63.9	66.2	68.7	74.4
% entered EBacc - EVER6	-	-	-	47.2	49.1	56.8
% entered EBacc – NonFSM	-	-	-	73.2	74.9	79.8
Odds Ratio	-	-	-	3.1	3.1	3.0
% achieved EBacc - EVER6	-	-	-	27.6	28.5	30.5
% achieved EBacc – NonFSM	-	-	-	55.5	56.6	58.4
Odds Ratio	-	-	-	3.3	3.3	3.2

Notes

(a) Includes double award although this only accounted for 0.3% of the cohort in each of 2013, 2014 and 2015.
Top three lines sourced from DFE SFR 01/2016. EVER6 and NonFSM averages are authors own calculations.

The percentage of student entering for the EBacc Science has increased substantially in the last three years, from 66.2% in 2013 to 74.4% in 2015, mostly reflecting an increase in entry to core and additional science with a slight decrease in entry for three individual sciences. However this has not been associated with any appreciable change in the EVER6 gap. The odds of EVER6 students entering EBacc science remain about one-third those of NonFSM students and their odds of achieving EBacc science are also one-third lower compared to NonFSM students.

This is not of itself surprising since achievement gaps related to poverty are large and long standing. Figure 2.1 plots the breakdown by entitlement to FSM of the headline 5 or more GCSE A*-C grades or equivalent including English and Mathematics (5EM) performance measure since its introduction in 2004. Despite substantial increases in achievement, including a doubling of the proportion of FSM pupils achieving the threshold, the relative gap as indicated by the OR has proved stubbornly large. While decreasing from OR= 4.0 in 2004 to OR= 2.9 in 2012, it has subsequently risen again to OR= 3.1 in 2015 (see Table 2.5).

Figure 2.1 Percentage of pupils achieving 5+A*-C including English and mathematics by entitlement to FSM 2004-2015
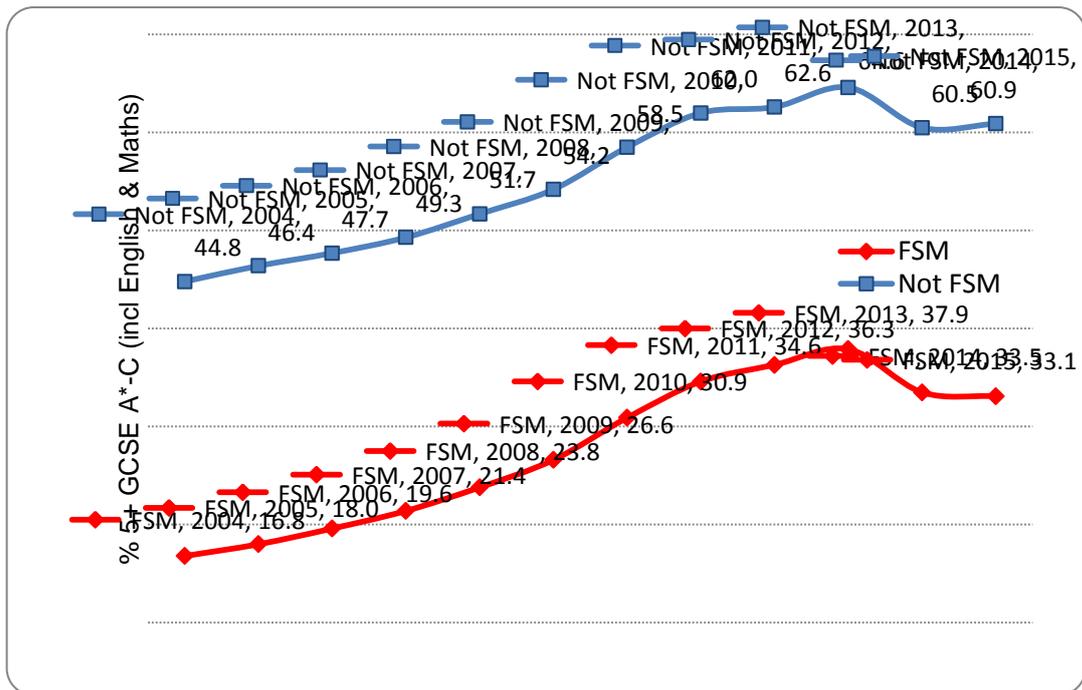
For comparative purposes Table 5 presents the gender and ethnic achievement gaps, contrasting for the latter the majority White British group with the lowest performing of the larger ethnic groups, Black Caribbean students. Significant gaps exist for these characteristics too, but the OR for ethnicity (OR=1.6) and Gender (OR=1.5) are about half the size of the OR for the FSM gap (OR=3.1).

Table 2. 5 FSM, ethnic and gender gaps in the percentage of students achieving 5 or more GCSE A*-C including English and mathematics 2004-2015

Student grouping	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
Ethnicity												
White British	40.9	42.9	44.2	45.8	48.0	50.9	55.0	58.2	58.9	60.5	56.4	57.1
Black Caribbean	22.8	27.1	29.2	32.7	35.9	39.4	43.5	48.6	49.8	53.3	47.0	45.9
Gap (% points)	18.1	15.8	15.0	13.1	12.1	11.5	11.5	9.6	9.1	7.2	9.4	11.2
Odds Ratio (OR)	2.34	2.02	1.92	1.74	1.65	1.59	1.59	1.47	1.44	1.34	1.46	1.57
Socio-economic												
FSM	16.8	18.0	19.6	21.4	23.8	26.6	30.9	34.6	36.3	37.9	33.5	33.1
Not FSM	44.8	46.4	47.7	49.3	51.7	54.2	58.5	62.0	62.6	64.6	60.5	60.9
Gap (% points)	28.0	28.4	28.1	27.9	27.9	27.6	27.6	27.4	26.3	26.7	27.0	27.8
Odds Ratio (OR)	4.0	3.9	3.7	3.6	3.4	3.3	3.2	3.1	2.9	3.0	3.0	3.1
Gender												
Boys	37.1	38.4	39.7	41.4	43.8	47.1	51.1	54.6	54.2	55.6	51.6	52.5
Girls	44.8	46.7	48.0	49.6	51.9	54.4	58.6	61.9	63.7	65.7	61.7	61.8
Gap (% points)	7.7	8.3	8.3	8.2	8.1	7.3	7.5	7.3	9.5	10.1	10.1	9.3
Odds Ratio (OR)	1.38	1.41	1.40	1.39	1.38	1.34	1.35	1.35	1.48	1.53	1.51	1.46

KS4 Breakdown by pupil and school characteristics

Appendix 2.4 presents a breakdown of the percentage of students entered for the EBacc science, and the mean EBacc two sciences points score, for a range of pupil and school background factors.

- Prior attainment band age 11
- Gender
- Ethnic groups (18 categories)
- English as an Additional Language (EAL) status
- Entitlement to FSM
- IDACI neighbourhood deprivation
- Special Educational Needs (SEN) Stage
- School Type
- School selective status
- Region

This data shows strong associations between all these pupil and school characteristics and EBacc science participation and achievement.

Much has been written recently about the London effect, so it is perhaps worth looking at the regional results in a little detail. Figure 2.2 below plots the EBacc science points score by IDACI deprivation decile[7] for each of the 9 regions of England. This does show substantially higher achievement in science in both Inner and Outer London compared to all the other regions of England, and particular so for students living in the more socio-economically deprived areas as indicated by the IDACI. For example among students living in the 10% most deprived neighbourhoods of England, those in London are on average scoring around 8-10 EBacc points above the other regions, or around one a half grades higher in each of their best two GCSE sciences.

However we need to be cautious because the capital differs from other parts of England in many ways. For example inner London contains 82% ethnic minority students compared to 20% in the rest of England, and ethnic minority students tend to be particularly highly achieving (Strand, 2014; 2015). Also while there is high economic deprivation in London the capital also contains areas of considerable affluence. If we restrict the analysis to White British EVER6 students we get a much smaller regional effect, as shown in Figure 2.3. London does not differ significantly from the East of England or the South West, although there is still a modest two point (as opposed to 10 point) advantage for London over the North East, North West, Yorkshire and East Midlands.

---

[7]. The Income Deprivation Affecting Children Index (IDACI) divides Lower Super Output Areas in England into 10 equally sized groups (deciles) each containing 10% of the England population, from the least disadvantaged 10% of neighbourhoods (Decile 1) through to the 10% most deprived (Decile 10).

Figure 2.2 EBacc science points score by IDACI deprivation decile and region 2015
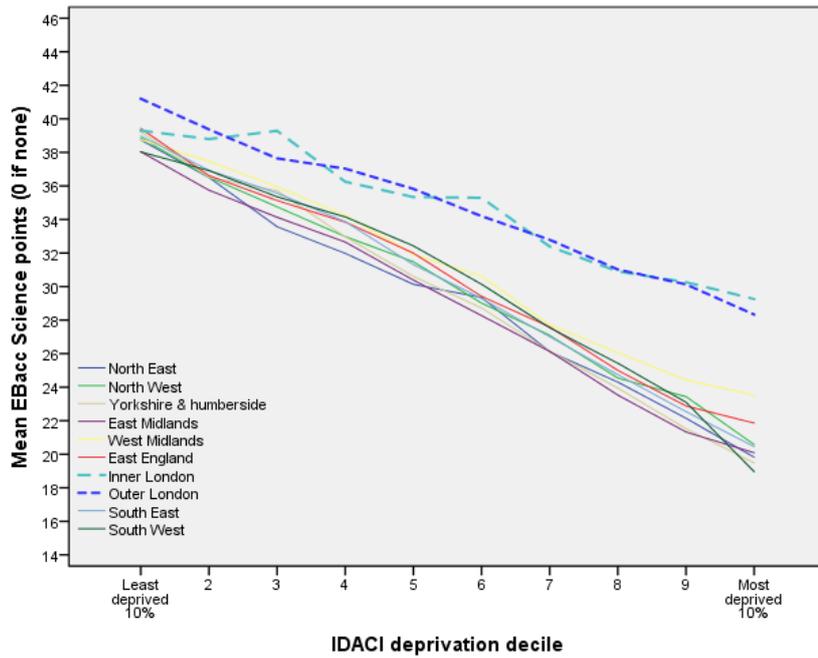
Figure 2.3 EBacc Science points score for White British EVER6 students by region
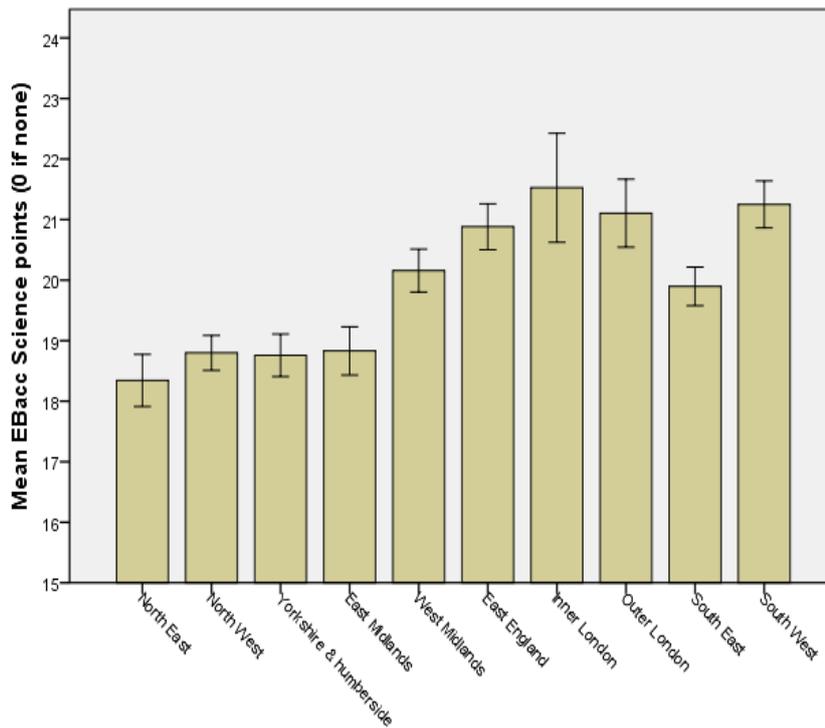
Further research might usefully interrogate the KS4 data further using multi-level modelling and other advanced statistical techniques. However at this point we move to analyse achievement gaps in science at KS5 (age 19).

Achievement at Key Stage 5 (age 19)

For technical reasons, described in detail in Appendix 2.5, we establish the achievement gap at age 19 by following a KS4 cohort through from age 16 to age 19. We are then able to give an accurate assessment of the EVER6 gap at age 19 from three bases:

- <u>KS4 cohort</u>: The total number of students in the cohort at the end of Y11. This base gives the direct proportion of EVER6 vs. NonFSM pupils at age 16 who go on to achieve Level 3 qualifications by age 19.
- <u>Level 3 cohort</u>: The number of students entered for Level 3 'A' level or vocational qualifications by age 19. This base gives the EVER6 vs. NonFSM gap at age 19 *conditional on having studied and entered for at least one level 3 qualification.*
- <u>'A' level cohort</u>: The number of students entered for at least one 'A' level or applied 'A' level qualification by age 19. This base gives the EVER6 vs. NonFSM gap at age 19 *conditional on having studied and entered for at least one 'A' Level.*

Through this analysis we can see how EVER6 status shapes whether students continue in education at all after age 16, the route they take (A level vs. Vocational Level 3 qualification route) and how these impact on their 'A' level achievement at age 19 and progress age 16-19.

Participation gaps at age 16-19

Table 2.6 presents the percentage of NonFSM and EVER6 students as recorded at the end of KS4 who go on to enter and achieve a range of outcomes at the end of KS5 (age 19). The percentage of NoFSM pupils, and the percentage of EVER6 pupils, achieving a range of outcomes are calculated separately for all pupils at age 16 (KS4 base); for only those pupils continuing in education and entering Level 3 qualifications (Lev3 base), and for only those pupils entering 'A' levels (Alev base).

The key points are:

- For pupils at the end of KS4, the odds for NonFSM students continuing to Level 3 qualifications are 2.60 times greater than for EVER6 students (67.8% vs. 44.8%). The contrast is particularly substantial for entering 'A' levels, where the odds for NonFSM are 3.1 times greater than for EVER6 pupils (45.1% vs. 21.0% respectively). A significant barrier to achievement in science is therefore participating in Level 3 study post 16.

- Among those that do enter for Level 3 qualifications, EVER6 students are more than twice as likely to be on the vocational route as NonFSM students (63.1% vs. 42.8%), and only half as likely to be on the A Level route (46.8% vs.66.6%). Therefore even among those who do continue into study for Level 3 qualifications, EVER6 students are skewed towards vocational study. Even when restricting comparison just to those who entered a Level 3 qualification, EVER6 students are around half as likely as NonFSM students to achieve 2+ 'A' levels or equivalent or three 3+ 'A' levels or equivalent.

- Among the A Level cohort, the odds for NonFSM students achieving 2+ or 3+ 'A' level passes are around twice as high as for EVER6 students, and even more disparate for the very highest levels of achievement (OR=2.7 for 3+ A*-A grades and OR=2.5 for AAB including two facilitating subjects).

Table 2. 6 Participation in KS5 including in Science subjects by EVER6 status at age 16

Variable	% of Never FSM			% of EVER6			Odds-Ratio		
	KS4 base	Lev3 base	A levl base	KS4 base	Lev3 base	A levl base	KS4 base	Lev3 base	Alev base
Post-16 pathway[a]									
Any Level 3 qualification	67.8	-	-	44.8	-	-	2.6	1.8	-
A' Level route	45.1	66.6	-	21.0	46.8	-	3.1	2.3	-
Vocational route	29.0	42.8	-	28.3	63.1	-	1.0	0.4	-
Level 3 achievement[b]									
Achieved 2+ A levels or equiv.	62.6	92.4	-	37.8	84.4	-	2.8	2.2	-
Achieved 3+ A levels or equiv.	56.0	82.6	-	31.6	70.5	-	2.8	2.0	-
Achieved 3+ A levels A*/A or equiv.	7.7	11.4	-	3.1	6.8	-	2.7	1.8	-
Overall 'A' level achievement									
Achieved 2+ A levels (A*-E)	40.7	60.1	90.2	17.4	38.8	82.9	3.3	2.4	1.9
Achieved 3+ A levels (A*-E)	34.1	50.4	75.6	13.2	29.5	63.1	3.4	2.4	1.8
Achieved 3+ A levels (A*-A)	4.6	6.8	10.2	0.9	1.9	4.1	5.6	3.7	2.7
Achieved AAB (incl 2 facil. subjs)[c]	5.8	8.6	12.9	1.2	2.6	5.6	5.2	3.5	2.5
Science 'A' level achievement									
Entered any science[d]	21.9	32.3	48.5	8.9	20.0	42.7	2.9	1.9	1.3
Entered an Ebacc science[e]	14.9	22.1	33.1	5.4	12.1	25.9	3.1	2.1	1.4
Biology	9.0	13.3	20.0	3.5	7.8	16.6	2.8	1.8	1.3
Chemistry	7.4	10.9	16.3	3.1	6.9	14.7	2.5	1.7	1.1
Physics	5.2	7.7	11.5	1.6	3.5	7.4	3.5	2.3	1.6
Electronics	0.2	0.2	0.3	0.1	0.1	0.3	2.8	1.9	1.3
Environmental science	0.2	0.3	0.4	0.0	0.1	0.2	3.8	2.5	1.8
Geology	0.4	0.6	0.8	0.1	0.2	0.4	5.1	3.4	2.4
Psychology	9.0	13.3	19.9	4.4	9.9	21.2	2.1	1.4	0.9
Computing	0.7	1.1	1.6	0.3	0.6	1.2	2.7	1.8	1.3
Applied Science	0.5	0.7	1.1	0.3	0.6	1.2	1.9	1.3	0.9
Number of students in KS4 cohort	416538	282345	188032	134795	60358	28241			

Notes.

(a) A level cohort entered at least one 'A' or Applied A level. Vocational Cohort entered at least one Level 3 (advanced) vocational qualification at least the size of an 'A' Level (i.e. 180 guided learning hours per year). Student can enter both 'A' level and Vocational subjects and therefore be included in both the A level and the Vocational cohort. (b) Around one-third of students achieve the 2+ and 3+ 'A' level or equivalents thresholds through equivalent qualifications (BTEC/OCR, General Applied, Technical Levels) so these percentages are only calculated for the whole Level 3 cohort. (c) Including two facilitating 'A' level subjects. (d) Any science includes all those subject defined as Science in the NPD (from Biology to Applied Science in the above table). (e) EBacc sciences are Biology, Chemistry, Physics and Computing.

- For the science subjects, the EVER6 participation gaps are largest when based on the whole KS4 cohort. For example 8.9% of EVER6 pupils enter for any science compared to 21.9% of NonFSM pupils (OR=2.9) and only 5.4% EVER6 pupils enter for one or more EBacc Science subjects compared to 14.9% of NonFSM pupils (OR=3.07). These gaps decrease somewhat when looking at just those students who enter for Level 3 qualifications, reflecting the more selective nature of the cohort, and even further when just looking at the gap for those who do enter 'A' levels. However even among the A Level cohort, NonFSM students are still more likely to enter a science subject than EVER6 pupils, particularly for EBacc Science (33.1% vs. 25.9%; OR=1.42).

- In terms of individual science subjects, and focussing on the 'A' Level cohort, participation gaps are largest in Geology (OR=2.37) and Environmental Science (OR=1.78), though the small number of entries indicate opportunities to study these subjects are rare. Among the more substantial subjects the participation gaps are largest for physics (OR=1.62) and maths (OR=1.40), and much smaller for Biology (OR=1.26), Computing (OR=1.27) and Chemistry (OR=1.14). There is no participation gap (conditional on entering for 'A' Level) for Psychology (OR=0.92).

'A' level points score gaps at age 19

Table 2.7 focusses on achievement as indicated by points scores for the 'A' Level cohort.

Overall achievement

For the 'A' level cohort, the EVER6 gap is 106 points, which is around three and a half 'A' Level grades. This is a large difference, in standardised terms Cohen's D= -0.41. Partly this reflects the fact that EVER6 students enter fewer subjects than NonFSM, 3.4 for EVER6 vs. 3.1 for NonFSM, so the difference in average grade per entry is somewhat smaller but still substantial (Cohen's D= 0.32).

Science achievement

Two summary measures of science achievement were created: (a) the average points score in science for students who entered any 'A' level science subject from Biology to Applied Science (as listed in Table 2.7), and; (b) a similar measure but restricted to just the four EBacc sciences (Biology, Physics, Chemistry and Computer Studies).

In total 103,000 students entered for at least one science subject, representing 18.7% of the KS4 cohort, and 47.8% of the 'A' level cohort. There was a sizeable EVER6 gap (Cohen's D=0.30) indicating poorer performance among the EVER6 students.

A smaller proportion of students entered at least one EBacc science subject, just 69,500 students, representing 12.6% of the KS4 cohort and 32.2% of the 'A' level cohort. Again EVER6 students achieved a much lower points score than NonFSM students with a Cohen's D of 0.32.

These two gaps are about the same size. Thus the EVER6 achievement gap is no bigger in the EBacc sciences than it is for all science subjects.

Looking at the individual 'A' level science subjects, for most subjects the achievement gap is about five points or half an 'A' level grade lower on average for EVER6 students (around one-third of a standard deviation or D= -0.33). This does not vary greatly across subjects,

though the gap tends to be a little smaller in Applied Science (d= -0.18), Environmental Studies (d= -0.18) and Computing (d= -0.15).

Table 2. 7 'A' level achievement and progress during Key Stage 5 by EVER6 status at age 16

Variables	Never entitled FSM			EVER6			Coh-en D
	Mean	N	SD	Mean	N	SD	
Overall achievement							
Total 'A' Level points score	**724.6**	188032	257.4	**617.8**	28241	246.5	0.41
Total A level entries	**3.4**	188032	0.83	**3.1**	28241	0.89	0.36
Average points per 'A' level entry	**209.3**	188032	42.9	**195.5**	28241	43.4	0.32
Science achievement							
Avg. A level points (any science/s)	**33.1**	91258	14.3	**28.8**	12049	13.6	0.30
Avg. A level points (EBacc science/s)	**33.5**	62258	14.8	**28.8**	7309	14.2	0.32
Mathematics	**39.5**	48942	15.2	**34.2**	5682	15.8	0.35
Biology	**34.6**	37676	15.2	**29.9**	4699	14.4	0.31
Chemistry	**36.7**	30686	14.5	**31.6**	4140	14.4	0.36
Physics	**34.7**	21613	15.7	**29.1**	2091	15.3	0.36
Electronics	**38.7**	640	15.4	**33.1**	74	16.2	0.36
Environmental science	**28.2**	767	14.1	**25.7**	65	14.1	0.18
Geology	**35.8**	1584	13.9	**31.6**	101	13.9	0.30
Psychology	**32.5**	37421	13.7	**29.2**	5993	13.4	0.24
Computing	**29.7**	2966	14.3	**27.5**	353	14.3	0.15
Applied Science	**24.9**	2044	10.8	**22.8**	341	10.8	0.19
Value added progress 16-19							
Average KS4 prior attainment points	**44.6**	282345	6.2	**41.5**	60358	6.6	0.49
Total 'A' level points - Value Added	**4.6**	188032	191.8	**-33.0**	28241	200.1	0.19
EBacc Science points - Value Added	**0.05**	62258	11.3	**-0.82**	7309	12.0	0.08

Notes. (a). EBacc Sciences are Biology, Physics, Chemistry and Computing. (b) For the value added EBacc science measure A* is awarded 60 points through to U awarded 0 points. (c) The denominator variables (KS5_TOTENTS) counts the size not just the number of entries i.e. a BTEC Diploma would count as three entries. (d) GCSE grades are scored as A*=58; A=52, B=48 through to G=16 and U=0. The total points are divided by the number of entries to calculate the average grade.

Value added or pupil progress 16-19

The last three rows of the table relate to value added calculations to show how much progress students make between the end of KS4 and the end of KS5. In terms of average prior attainment score at KS4, those students who were EVER6 score around half a SD lower than NonFSM, so there was a large difference in achievement before students started their A Levels. The value added scores are the result of a simple linear regression to show how the total 'A' level points score and the EBacc science points score change when a control for prior attainment at KS4 is included.

- The EVER6 achievement gap for 'A' level total score 'value added' score is 37 points, or just over one 'A' level grade. Thus although the EVER6 gap more than halves from D=0.41 for the raw score to d= 0.19 for the value added score, EVER6 students are still making less progress than their NonFSM peers during the age 16-19 phase.

- For the 'A' level EBacc science subjects score, while the EVER6 gap in the 'raw' measure was 4.7 points or almost half an 'A' level grade (d= 0.32), the value added gap is only 1 point or one-tenth of an 'A' level grade (d=0.08). We can conclude that those EVER6 students who choose to study EBacc sciences make only slightly less progress than their NonFSM peers, and that around 75% of the achievement gap in 'A' Level EBacc sciences can be attributed to low prior attainment. The main issue for the EBacc sciences seems to be low prior attainment combined with the small number of EVER6 students who choose to study 'A' level EBacc science (as we saw in Table 2.6, 15% of NonFSM pupils but just 5% of EVER6 pupils take an 'A' level in an EBacc science).

Conclusions from analyses of the NPD

Table 2.8 below attempts to summarise some of the main conclusions arising from the NPD analyses by drawing some key statistics from the preceding tables. Any attempt to analyse such a large and varied set of data are obviously fraught with complications and qualifications, but we attempt to draw some general conclusions below.

- EVER6 achievement gaps are evident from the earliest point at which national achievement data is collected, namely the end of Reception Year when children are aged 5. Children who are EVER6 score half a standard deviation below the mean for NonFSM pupils in the Early Years Foundation Stage Profile (EYFSP) and the odds of their demonstrating a good level of development (GLD) are 2.3 lower than for NonFSM students.

- Science achievement gaps are first measured at age 7 at the end of KS1. They are large, with EVER6 pupils scoring 0.43 SD below NonFSM pupils and the odds of achieving the expected level 2 or above 2.3 times lower than for NonFSM pupils. Low achievement in Science is strongly predicted by EYFSP score at age 5 and it is likely that science achievement gaps reflect these early learning gaps (see further discussion below).

- EVER6 achievement gaps in science remain throughout compulsory schooling, growing slightly larger by the end of secondary school. Thus at age 16 EVER6 students are three times less likely to enter for the EBacc in Science and score 0.63 of a SD below NonFSM students in EBacc science points score.

- EVER6 children make less progress in Science than NonFSM at every key stage, meaning achievement gaps tend to widen over time. However it seems the gaps grow particularly strongly between ages 5-7 and age 11-16. These might be areas of particular focus for intervention, though the focus of intervention might be quite different in the two phases (see further discussion below).

Table 2. 8 Summary of standardised EVER6 achievement gap measures age 5-19

Key stage	age	Science				All subjects average			
		Measure (D/OR)	D	OR	VA	Measure (D/OR)	D	OR	VA
Reception	5	-	-	-	-	EYFSP score / Good Level of Development	0.50	2.3	-
KS1	7	Science TA / Level 2+	0.43	2.6	0.26	KS1 Average points score / Level 2+ all subjects	0.51	2.6	0.21
KS2	11	Science TA / Level 4+	0.46	2.7	0.10	KS2 Average points score / Level 4+ all subjects	0.49	2.3	0.09
KS4	16	EBacc Points Score / Entered EBacc Science	0.63	3.0	0.26	Best 8 points score / 5+ A*-C incl. EM	0.71	3.1	0.32
KS5 (all KS4 students)	19	Entered any 'A' level Science	-	2.9	-	2+ A level or equiv. (Level 3) passes	-	2.8	-
KS5 (Level 3 cohort)	19	Entered any 'A' level science	-	1.9	-	2+ A level or equiv. (Level 3) passes	-	2.2	-
KS5 (A Level cohort)	19	Avg. A level points (Science) /Entered any Science	0.30	1.3	0.08	Total A Level points / 2+ A level passes	0.41	1.9	0.19

Notes
Measure indicates the specific measure employed to calculate Cohen's D (D) and the Odds Ratio (OR) respectively. VA= Value added measure of progress during the key stage. Data drawn from 2015 NPD for England. See previous tables in this chapter for detailed description of data sources and measures.

- After age 16 data are no longer directly comparable to earlier phases because full-time education becomes optional and there is no requirement to continue to Level 3 study. Among the set of students who continue to study and enter at least one 'A' Level science there is still an achievement gap, with EVER6 students achieve an average 'A' Level science 0.30 SD lower than their NonFSM peers. While this is smaller than Cohen's D in earlier phases this reflects the differential continuation post 16. For example 55% of EVER6 students compared to just 32% of NonFSM students do not go on to take any Level 3 qualifications by age 19. In science only 8.9% of the EVER6 students at the end of KS4 enter a Science GCSE, compared to 21.9% of NonFSM (OR=2.9). The biggest gaps in Science achievement post 16 are therefore due to non-participation.

- The general pattern of achievement gaps and progress scores for science are broadly similar to the achievement gaps and progress scores for the overall measures of achievement at each key stage.

These results chime with other recent research in the area of science achievement. For example Morgan, Farkas, Hillemeir and Maczuga (2016) report a representative longitudinal

sample of 7,757 children from the US Early Childhood Longitudinal Study Kindergarten followed from age 5 to age 14 between 1998 and 2007. They conclude that kindergarten general knowledge was the strongest predictor of first grade general knowledge which was in turn the strongest predictor of children's science achievement from third to eighth grade. They conclude that efforts to address science achievement gaps in the US are likely to require intensified early intervention efforts, particularly those delivered in the early years. For example we know that low SES children on average have lower access to high quality child care and pre-schools, have fewer educational resources at home, have less stimulating home learning environments, are more likely to struggle with reading and mathematics needed to access and develop scientific concepts, vocabulary general knowledge and so on (Sylva, 2014; Strand, 2014).

However opportunities to learn during school are also likely to be significant, and this may be particularly important in the early years of primary school as well as during secondary school. The importance of attention to science in primary school has recently been emphasised by the Wellcome Trust (2014) and to this we would add a particular focus on Foundation stage and Key Stage 1. Regarding secondary schools, it is clear the most significant predictor of entry to EBacc science is prior attainment, but there does also seem to be variation related to factors like region of England even when holding ethnicity and deprivation constant. Further research might usefully interrogate this data further using multi-level modelling and other advanced statistical techniques.

Chapter 3

Exploring the cause of any SES related attainment gap in science: SES, attainment and interest in science[8]

Box 3 Chapter 3 Summary

This chapter is based on a systematic review and reports on the robustness of the SES-science attainment gap as well as on quantitative analyses of large data sets that examined possible mediators of this gap.

- The robustness of the SES-science attainment relation.

    1. The effect of SES on science attainment is robust: it has been replicated in a large number of studies since 1971 in different countries and using different measures of SES and of science attainment.
    2. It has been found in more affluent as well as less affluent countries.
    3. It has been found from the end of kindergarten and remains throughout primary school.
    4. In the PISA 2015 results for the UK, SES explained 10.5% of the variance in science attainment; the difference between students in the top and in the bottom quartile was 84 points (the mean for the UK was 509 points); 35% of pupils in the lower SES level perform similarly to those in the higher SES level (an increase of 5% in comparison to 2006).

- The search for mediators of the effects of SES on science learning in the papers that report quantitative analyses of large data sets produced two hypotheses: (1) that this effect is mediated by an opportunity gap; (2) that this effect is mediated by an interest gap.

- The concept of opportunity gap has been used in the literature to label the effect of differences in resources for learning science on attainment.

    1. At the country level, it has been found that children who attend schools in low income countries learn less science than those who attend schools in high income countries after being in school for comparable periods of time.
    2. At the school level, opportunity differences are also noted. The school SES level, which is the average SES of their pupils, explains variation in science attainment scores above and beyond individual SES. In some countries, teacher qualification and level of resources allocated to science teaching (e.g. materials for laboratory and time dedicated to science teaching) correlate with differences in the school SES levels. SES accounts for a larger proportion of the variance in science attainment in educational systems which include tracking (i.e. allocation of students to different types of education during secondary school) than in those systems which do not.
    3. At the family level, pupils from lower SES backgrounds have access to fewer educational resources (e.g. desks, dictionaries). The differences in individual pupils' SES account for differences in their science attainment, even after the effects of school differences in SES have been taken into account.

- Is there an SES gap in interest in science?

    1. Attainment differences in science learning do not appear to be mediated by differences in motivation and interest.

---

[8] This chapter was prepared by Terezinha Nunes, Peter Bryant, & Rossana Barros

2. At the country level, participants in the higher scoring countries show less interest in science, irrespectively of how it is measured.
3. Within countries, there is a positive but only a modest relation between interest and attainment in science. Across studies, there is no consistent evidence for a negative correlation between SES and interest in science.

Our aim in this and the two subsequent chapters is to answer the question: what are the mediators of the relation between SES and science attainment? (see Appendix 1.5 for a definition of mediators and moderators) Our approach in searching for an answer involved two different searches of the literature and a set of analyses of the longitudinal data from the Avon Longitudinal Study of Parents and Children (ALSPAC). In this chapter, we report the outcomes of the first search, which aimed to identify research that included data both on SES and on science learning, as well as an explicit hypothesis about what might cause the association between SES and science learning. Chapter 4 reports on the second literature search, through which we aimed to relate the conditions that lead to good attainment in science learning and SES, but in this second search we no longer required all three variables - SES, the possible mediator, and science attainment - to be measured and analysed in the same study. Chapter 5 reports analyses of data from ALSPAC in order to test hypotheses about plausible mediators identified in Chapter 4.

In the first section of this chapter we ascertain whether the relation between SES and attainment in science is indeed robust. The second section considers the evidence for the hypothesis that the link between SES and science attainment is explained by an opportunity gap: children from poorer backgrounds simply do not have as much opportunity to learn science as those from more prosperous backgrounds. The third section analyses the evidence for the hypothesis that children from poorer backgrounds do not attain as much in science simply because they are not interested; interest in science is the measure of motivation to learn science used in the vast majority of studies that consider this affective aspect of learning.

SES and Science Attainment: Is There a Robust Connection?

SES effects on children's performance in science measures have been investigated using different methodologies, which depended on the specific research questions that they addressed and on the research designs and statistical methods available at the time of the study. The earliest study identified in our search that fits the criterion of collecting and analysing appropriately information on SES and student attainment was by Klein (1971), who identified 15 primary schools in Minnesota that recruited children from homogenous SES backgrounds, five of which were attended by children from lower SES, five by children from middle SES and five from higher SES families. The children (N=305) participated in an assessment of science concepts and also answered questions about how they might find out the answer to the question, if they did not know it. The science concepts test showed a good level of reliability and was validated by a significant correlation with the outcomes of an interview, during which the children's understanding of the concepts was probed. Results with children from ethnic minorities were excluded from the analysis in order to avoid confounding SES and minority status. A month before the science assessment, the children were given the Lorge-Thorndike intelligence test. An analysis of covariance, which took into

account the differences in the intelligence test scores, showed that there were significant differences between the low SES children and the other two groups in the number of correct responses to the concepts test. The children from the low SES group were also less likely to prefer experimentation as a method of finding out answers to science questions than the children from the other two groups. Klein's use of a statistical control for performance in an intelligence measure allows us to reject the hypothesis that SES and science outcomes are related only because they are both consequence of a single cause, intelligence. In her literature review, Klein noted that she was not able to find previous statistical analyses of the relation between SES and science attainment and attributed this gap in the literature to the previous lack of suitable statistical methods for pursuing such investigations.

Since Klein's pioneering study, many other studies investigating the relation between SES and science attainment have been carried out, the large majority of which uses results of high stakes tests such as the US based National Assessment of Educational Progress (NAEP) as well as international data sets such as the International Association for the Evaluation of Educational Achievement (IAE), the Second International Science Study, the Trends in International Mathematics and Science Study (TIMSS) and the Program for International Student Assessment (PISA). A high proportion of the studies uses Hierarchical Linear Models. Some studies include a diversity of predictors in the equations, which allow for exploring whether the relation between SES and science attainment can be seen as specific (in the sense that it does not disappear if other predictors are included in the equation) or whether it could be a product of the association between SES and some other factors. In spite of variations in the measures of SES and science outcomes, in the statistical techniques employed and in the predictors included in the equations besides SES, the relation between SES and science attainment stands as robust: 16 studies (Anderson, Lin, Treagust, Ross, & Yore, 2007; Areepattamannil, Shaljan, & Kaur, Berinderjeet, 2013; Baker, Goesling, & LeTendre, 2002; Beese, & Liang, 2010; Gilleece, Cosgrove, & Sofroniou, 2010; Ho, 2010; Maerten-Rivera, Myers, Lee, & Penfield, 2010; Kohlhass, Lin, & Chu, 2010a and b; Lin & Shi, 2004; Lynch, Benjamin, Chapman, Holmes, McCammon, Smith, & Symons, 1978; Marks, 2006; McConney, & Perry, 2010; Perry & McConney, 2010; Yang, 2003; Young, 1990) in diverse regions (e.g. Australia, Canada, China, Finland, Hong Kong, Ireland, Korea, New Zealand, Singapore, Taiwan, USA) and excluding participants with ethnic minority or immigrant status, report statistically significant effects of SES on science attainment in different outcome measures.

A recent analysis of a large scale longitudinal study by Morgan, Farkas, Hillemeier, and Maczuga (2016) in the US shows that the SES effect on school attainment appears as early as the end of kindergarten and persists throughout primary school. Morgan, Farkas, Hillemeier, and Maczuga divided the participants into quintiles by SES-level; the gap between the quintiles remained practically the same from 3rd to 8th grade. With each increase by 1 standard deviation in the SES scale, scores in the science measure went up by .3 standard deviation[9].

---

[9] The standard deviation is a measure of the variability in a data set. It is calculated by finding the difference between each score and the mean, squaring this difference (so that there are no negative values), and finding the mean of the differences. In a normal curve, the distribution of the scores tends to be within a total of six

The study by Baker et al. (2002) provides the opportunity for contextualising the SES effect on science outcomes in the UK, as it included pupils in 36 different countries. Family SES was measured by a composite score based on three indicators: the mother's and the father's educational level, and number of books in the home. The outcome measure was the 1994-5 TIMSS science data. The amount of variance explained by SES in the science measure varied between 1.5% (Taiwan) and 19.2% (UK); other countries where the amount of variance explained by family SES was large were Hungary (17.6%), Germany (16.6%), Switzerland (13.9%), Portugal (12.5%) and Singapore (12.4%).

Previous PISA results showed that 23% of the variance in science scores in the UK was explained by SES (Marks, 2006), but PISA 2015 results show a more positive outcome for the UK: 10.5% of the variance in attainment was explained by SES.

Conclusions about the robustness of the link between SES and science attainment

The evidence for the existence of SES effects on science attainment is robust. The results include participants aged between 6 and 15 years and cover a span of about four decades; the PISA results are confirmed in Australia for 2003 and 2006 assessments (McConney, & Perry, 2010; Perry & McConney, 2010). The persistence of the gap has been documented between the end of kindergarten and 8th grade in the US. These are all large scale studies, in which the smallest number of participants is higher than 1,400 and many include more than 4,000. However, none of the PISA and TIMSS datasets in these 16 studies included measures of possible mediators and longitudinal analyses of these mediators as predictors of science attainment. The exception to this is the study by Morgan, Farkas, Hillemeier, and Maczuga (2016), which is described in Chapter 4 in greater detail.

Explaining the SES gap in science attainment

Baker et al. (2002) suggest that, by the late 1960s, the impact of family SES on education outcomes were clearly established (for the UK, see Plowden, 1967) and the search for explanations and the identification of malleable factors took the centre stage. According to Lareau (1987), previous research focused mostly on outcomes, but attention then turned to the processes through which these educational patterns were created and reproduced (e.g. in the UK: Bernstein, 1975; Rutter, Maughan, Mortimore, & Ouston, 1979). Different (not necessarily alternative) hypotheses emerged to explain the impact of SES on educational outcomes. Some researchers (mostly anthropologists, sociologists and psychologists) focused on the interactions that took place in the classroom as well as the relationships between parents and schools. Others sought to analyse quantitative data at institutional level. In this chapter, the focus is on the institutions: the country, the school, and the family. Chapters 4 and 5 analyse possible mediators at the individual level.

We have grouped the analyses at the institutional level in two themes: 1. The SES gap and the opportunity gap; 2. SES and motivation: is there an interest gap?

---

standard deviations. One can think of an increase in .3 of a standard deviation as moving up one step out of a total of 20 possible steps.

The SES gap and the opportunity gap

In the studies reported here, three types of institution are considered: the country, the school, and the family. Many of the studies rely on the large data sets from TIMSS and PISA, but other international studies and national studies are also included.

a.   The opportunity gap at the country and the school level

The collection of large amounts of quantitative data from different countries was initiated by the International Association for the Evaluation of Educational Achievement (IAE), housed in UNESCO's Institute of Education. The world was conceived as a big laboratory in which countries could test out their curricula and the effectiveness of their school systems (Comber & Keeves, 1973). The focus of the international comparisons was therefore on the institutional level analysis, with the aim of identifying variables that could be targeted by policies.

In a pioneering study, Comber and Keeves' (1973) investigated the effect of home background and school factors in a study which included data on 10,000 schools, 50,000 teachers, and 260,000 pupils in the age range 10 to 14 years in 18 countries. It included approximately 500 variables; SES, school type and teacher quality were among the predictors of science scores. The strategy for the analysis was to average the results across the 18 countries and to include those variables that had a beta coefficient of at least .05 as predictors of scores[10]. This means that variables that had a small impact on science scores, when all the other variables where also being taken into account, were excluded from further analysis; all variable mentioned in here met this criterion. There were large country differences in science attainment, which correlated with the country's wealth (measured by per capita income). The analysis also showed that the combined measure of parental education, occupational prestige, and other indicators of home circumstances had a larger effect on children's science attainment than the sum total of the influence of school and teacher quality as measured in the study. They concluded that SES affected the children's science attainment more than school factors, including teacher quality.

The conclusions from this well-known study were challenged subsequently, when Heyneman and Loxley (1982; 1983; see also Heyneman, 1976) re-analysed this data set, considering the fact that there are correlations between a country's wealth and science attainment. In their view, previous analyses of the SES impact on educational outcomes had been carried out in a limited number of countries (mostly in Europe, North America, and Japan) with well developed school systems and using methods that averaged effects across countries, leaving no room for the consideration of effects that were important for some but not for other countries. This criticism was also made about the Plowden report, which stated that the amount of variance in academic attainment accounted for by children's experiences prior to entering school considerably exceeded the impact of school quality on

---

[10] Beta coefficients indicate the direction and the strength of a the relation between a predictor and an outcome variable in a multiple regression. It indicates the number of standard deviations that the outcome measure will change as a result of one standard deviation in the predictor. For example, a beta coefficient that is equal to .05, as the criterion chosen by Comber and Keeves, indicates that when the predictor increases by one standard deviation, the level of science attainment increases by 5/100 of a standard deviation in that regression analysis. Because the beta coefficient reflects the influence of all predictor variables in a multiple regression model, it is different from a correlation, which only considers the association between two variables.

educational outcomes. Heyneman and Loxley (1982; 1983) implemented some changes to the statistical procedures used in the studies and included a greater variety of educational systems in their analysis, such as Uganda, El Salvador, Brazil, Paraguay, Mexico, Peru, Colombia, Argentina, Bolivia, Egypt and Botswana. This new analysis produced results that seem judicious and clear. First, children who attended schools in low income countries learned much less science than those who attended schools in high income countries after being in school for comparable periods of time. Second, the school effect on outcomes for children from low income countries was comparatively greater. In high income countries the school quality variables explained 11.4% of the variance in science attainment; in low income countries, school quality explained 20.8%. In brief: better school systems produce better outcomes, and the effects of school quality is more easily detected in low income countries, where schools vary greatly in quality.

Considering that poorer countries have made large investments in education since these initial results, Baker et al. (2002) examined whether the same trends could be found in the TIMSS data for 1994-95. They used the same statistical approach as Heyneman and Loxley and looked at variables that measured very similar aspects of the home environment and school quality. The relationship between a country's wealth and science attainment was again positive and significant, even though it had reduced in intensity, but it was no longer true that school effects were stronger in poorer nations. Baker et al. interpret this change as a consequence of the expansion in quantity and quality of school systems in poorer nations; they consider that, after this expansion, it is more appropriate to think of a "continuum running from dominant school effects to dominant family effects on attainment, varying by the degree to which poor nations incorporate minimum standards of school quality throughout the nation" (Baker et al., 2002, p. 307). In the UK, individual SES continued to be significant in the multilevel analysis, after controlling for the country's per capita income and school resources; the total variance explained by individual SES was 19%.

The recent data from PISA 2015 showed that 10.5% of the variance in science scores in the UK was explained by individual SES (Mo, 2016). This is a considerable reduction when compared to previous results, when SES explained 23% of the variance in science attainment in the UK (Marks, 2006). Another indicator that suggests that the effect of SES on science achievement in the UK has improved is the percentage of resilient students - i.e. students who are at the bottom 25% in SES but show an attainment comparable to those at the top in SES. In the 2015 PISA data, 35% of disadvantaged students in the UK were resilient; this is an increase of five percentage points from 2006. It is tempting to relate this positive trend to investments in science teaching and policy changes, but further data analyses would be required for this connection to be established.

PISA data from 2015 continues to show a variation between richer and poorer countries; the correlation between per capita GDP (Gross Domestic Product) and investment in education is high (r=0.91; OECD 2016 c; pp 185; OECD: Organisation for Economic Co-operation and Development); "36% of the variation in mean scores is associated with differences in per capita GDP across countries and 55% of the variation in mean scores is associated with differences in cumulative expenditure on students up to age 15" (OECD b, pp 265).

This result is in line with the conclusions drawn from the Heynemann and Loxley's analysis, which in essence reflects an opportunity gap between countries and within countries: richer nations invest more in education and their pupils gain better science scores than those in

poorer nations; some schools in some poorer nations do not offer their pupils the minimum standards in quality, and this is captured by school effects being stronger when rather weak schools are included in statistical analyses.

More interestingly, the results of PISA 2015 indicate that the relation between investment in education and attainment is not linear. "Among the countries and economies whose cumulative expenditure per student is under USD 50 000 (the level of spending in 18 countries), higher expenditure on education is significantly associated with higher PISA science scores. But this is not the case among countries and economies whose cumulative expenditure is greater than USD 50 000, which include most OECD countries" (OECD, 2015 b, pp 185). In the countries where the cumulative expenditure is less than USD 50 000, the amount of variance explained in scores is 41%; in those where the expenditure surpasses this limit, only 1% of the variance is explained by expenditure, and it matters more how the resources are spent. In brief, richer countries spend more on education than poorer countries, and their students attain higher science scores, but once a sufficiently large investment is made, the gains in score are no longer systematically related to the investments. It is therefore necessary to contextualise the opportunity gap by considering differences in how richer countries invest in science education.

Finally, there are also surprising results in the PISA 2015 analysis of the relation between SES and science attainment. The performance of students sharing similar socio-economic circumstances across countries and economies can vary widely. For instance, in Macao (China) and Viet Nam students facing the greatest disadvantage on an international scale have average scores of over 500 points in science, well above the OECD mean score. These disadvantaged students outperform the most advantaged students internationally in about 20 other PISA-participating countries and economies. The PISA 2015 report cannot offer any explanation for these unexpected findings, because in Macao only 1.7% of the variance in science attainment is explained by SES whereas in Viet Nam this percentage is 10.8, which is not radically different from the UK results. Such resilience underscores the fact that the SES effects in science attainment are no inevitable, and therefore must be mediated by intervening variables.

b.   Contextualising the opportunity gap

These general trends can be contextualised by considering case studies that describe the allocation of school resources in different countries. The first case study considered here is based on PISA 2006 science attainment and compares the US, Canada and Finland, whose mean science scores were 488.57 (US), 520.69 (Canada) and 563.59 (Finland). Beese and Liang (2010) point out that the US is a wealthy nation but has the highest proportion of children living in poverty of the three countries: 21.9% for the US, 14.9% for Canada, and 2.8% for Finland. The US also spent at the time least among the three countries per pupil per year: Finland spends US $6,440 per pupil whereas the US spends US $5,031; the figure for Canada was not reported. On PISA's equity measures, which take into account the percentage of disadvantaged pupils that show high attainment, the US ranked 45 of 55 countries. In the US, the allocation of school resources is determined by local community tax revenue; the wealthiest schools spend about 10 times more resources per pupil than the schools with the most disadvantaged populations whereas in Finland money is distributed centrally and equally (per pupil allocation to schools).

In all three countries (as elsewhere), there are differences between schools and also differences between students within schools. Because students within a school share the same school environment in so far as the resources are concerned, researchers assume that the differences between students within a school are more likely to be related to their home background, whereas differences between schools are more likely to be attributed to the resources that the schools assign to science teaching. When the between- and within-school differences in the science scores were separated, the between-schools differences in Finland explained 5.8% of the variance, which indicates greater homogeneity across schools, whereas in the US they explained 23.8% and in Canada, 20.5%. In Finland, there were no school resources measures that explained differences in science attainment between schools. In the US and Canada several variables describing the school resources were related to differences between schools in average attainment: school type (private versus public), shortage of lab equipment, shortage of science teachers, and ratio of full-time versus part-time science teachers were significantly related to science attainment. On average, students in private schools in the US scored 57.53 points higher in PISA than those in state supported schools; those in state supported schools where principals reported lack of lab equipment and lack of qualified teachers scored on average 10.12 and 2.69 points lower, respectively.

In summary, in Finland there were no resources differences between schools that were related to the students' attainment. In the US and Canada, resources differences between schools were related to differences in pupil attainment. In the US, individual SES differences were correlated with school differences in resources for teaching science, which were related to science attainment. This means that there was an opportunity gap associated with the SES gap.

The recent PISA results show that the US displayed the largest decrease in amount of variance explained by individual SES in science attainment from 2006 to 2015: the amount of variance explained by individual SES for science results in the US was 11% in 2015. The percentage of resilient students, who are at the bottom 25% in SES but show an attainment comparable to those at the top in SES, grew in the US from 25.0% to 31.6%. In the United Kingdom, 35% of disadvantaged students are resilient; this is an increase of five percentage points from 2006 (OECD, 2016).

The second case study is based on a large scale study conducted in China. Zhang and Campbell (2015) collected data from a national science assessment on almost 10,000 children in middle school (over 300 schools) and more than 2,000 science teachers in six provinces. Samples from different regions in China were collected to represent different levels of economic development; the eastern areas are well-developed and the western areas are under-developed. The individual SES index was composed by considering parent education level, parent occupation, and family properties; the individual SES was aggregated at the school level to describe the overall school-level SES.

The between school differences accounted for 26.8%, 25.7%, and 25.4 %, respectively, of the variance in biology, physics and earth and space assessments. School-level SES and teacher characteristics (teacher qualification, teaching experience, and a first degree in science in the same domain they were teaching) were correlated. After taking individual SES and gender into account (the study always uses these two variables together), teacher characteristics still had a significant impact on science attainment, but when school SES was

controlled for, teacher qualification was no longer significant. Because the school SES was highly correlated with teacher qualification (the higher the school's SES level, the more qualified were its teachers), this lack of teacher qualification effect does not mean that teacher qualification is unimportant. In fact, a more appropriate multi-level analysis would have been to test whether teacher quality mediated the school SES effect, but this analysis was not reported. In brief, individual and school SES explained significant amounts of variance in science attainment; the latter are highly correlated with teacher qualification, a result that illustrates the opportunity gap in China.

A third case study that illustrates the opportunity gap is provided by McConney and Perry (2010) and Perry and McConney (2010), who analysed PISA science scores in 2003 and 2006 in Australia. Using the same idea of disaggregating scores by SES groups, they divided the pupils into five SES groups (from high to low) and each of these five groups into quintiles by school SES (i.e. the average school SES), producing 25 groups for their analysis. The smallest group included 93 participants and the largest over 1,000; the total number of records was over 14,000.

McConney and Perry's (2010) strategy for data analysis provides a clearer picture of the individual and school level SES effects on science attainment than multilevel models because it allows for comparisons between students of the same SES who were attending schools with different school SES levels. Figure 3.1 summarises the results of the individual and school SES differences in science attainment in Australia for PISA 2006.

Figure 3. 1 Patterns of Science Attainment by Individual and School SES (based on data from McConney & Perry, 2010)
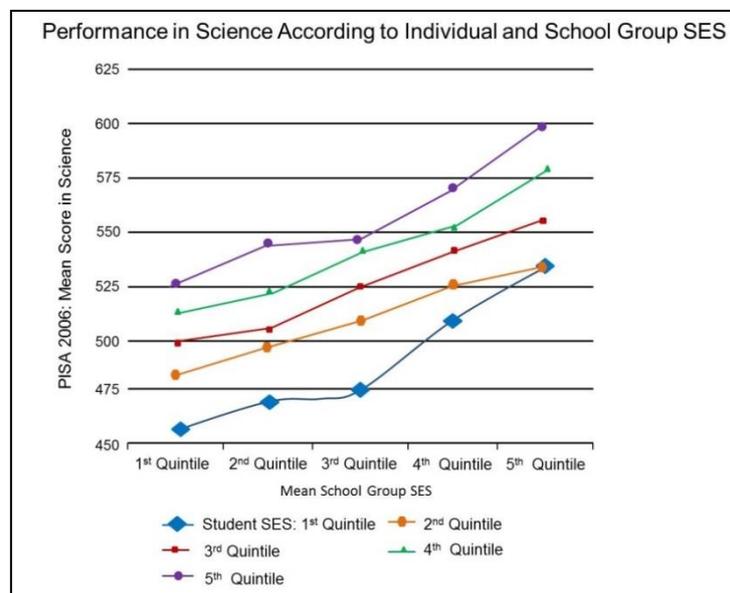
Substantial differences were found among students in the high SES and low SES schools for students with the same individual SES: for example, the average science score for a high SES student in a high SES school was 0.75 standard deviations higher than the average science score for a high SES student in a low SES school. In fact, the performance of a high SES

student in a low SES school was comparable to that of a low SES student in a high SES school.

A fourth case study is provided by Mere, Reiska and Smith (2006), who analysed the data for Estonia's performance in TIMSS 2003 for 8th-grade students. Estonia was a high performing country, at the top of all European counties in science attainment and the 5th in the full list of participating countries. Their analyses of SES impact on attainment showed significant effects by school and by individual SES. High SES classes were on average more advantaged than medium or low SES classes in instructional and organizational resources as well. Students in high SES schools were more likely to have teachers with a major in the subject taught, less likely to have teacher with less than three years of experience, and more likely to have teachers who reported being prepared to teach the topics covered on the TIMSS assessment. High SES students were also less likely to have teachers who reported that there were factors that limited their instruction and more likely to report a positive school climate. In their full multilevel model, school and individuals' SES were significant independent predictors of science attainment (i.e. they both explained variance in attainment when they were included in the same equation). Teachers' reports on whether or not there were factors that limited their opportunity to teach science were also significant in the same equation, explaining variance independently of individual and school SES. Other teacher variables (teacher background and less than three years' experience) were not significant when school SES was in the equation. As in the previous study, this does not mean that teacher quality does not matter because it is not possible to separate out the effects of teacher quality from school SES; it simply means that school SES effects are larger than teacher quality effects. Teacher quality may be part of the explanation of school effects but, because teacher quality and school SES are correlated, it does not explain variance independently of school SES effects.

School SES effects are confounded in some countries because in school systems children are assigned to different sorts of schools, such as grammar or comprehensive schools, before they start secondary education. In other countries they are not assigned to different types of school but there is the practice of streaming, which refers to the assignment of children to different classes expected to cater for children of different levels of ability in the same school. The research considered here treated tracking and streaming as a single factor, thus a single term, tracking, is used here. Tracking is definitely an institutionalised source of opportunity gap between students. Some researchers suggest that tracking is closely associated with SES (e.g. Alexander, & Cook, 1982; Jones, Vanfossen, & Ensminger, 1995), but Marks (2000) argues that the evidence for the association between tracking, SES and pupils' attainment is not clear. Essentially, his largely implicit argument is that tracking might be associated with SES because both are associated with ability. He analysed the PISA 2000 data for 15-year olds in 30 countries in order to describe the relative importance of SES and of tracking for science attainment. At the country level, he found larger SES effects on attainment when tracking was part of the system. On average, students' SES accounted for 23% variance of the effect of their placement in the tracking system on science outcomes. Marks concluded from this that SES accounts for a small reduction in differences in science attainment between students in different tracks in the education system; we note that the moderate correlation (r=0.53) between the two is often considered scientifically important in education, beyond its statistical significance. The results of the association between SES and student location in the tracking system for the UK were stronger than the

average results for the 30 countries: the school differences accounted for 24% of the variation in attainment; there was a reduction of 45% in the school effect when SES was controlled for.

The PISA 2015 report, selecting students into different programmes or schools, especially when students are young, is strongly associated with less academic inclusion across schools and less equity in science performance. On average across OECD countries, students in general programmes score 22 points higher on the PISA 2015 science assessment than those enrolled in pre-vocational or vocational programmes, after accounting for students' and schools' socio-economic profile. The most common practice across OECD countries is to select students into different tracks at age 15, but some countries start earlier; in 27 countries, including the UK, 99% or more of the participants were enrolled in general rather than vocational programmes.

The idea that variations in opportunities to learn about science at school might play an important part in pupils' attainment in science is incorporated in a hypothesis about mathematics and science learning, which was developed in a well-known study by Byrnes and Miller (2007). Their central idea was that attainment in mathematics and science is mainly a function of three factors: these are "(a) opportunity factors (e.g., coursework), (b) propensity factors (e.g., prerequisite skills, motivation), and (c) distal factors (e.g., SES)" (Byrnes & Miller, 2007 p. 599). To test this hypothesis, they analysed data from the National Educational Longitudinal Survey (NELS) on about 15,800 US children, gathered while they were in middle school and later on when they were in the 10th and the 12th grades at their secondary schools. This data bank contained detailed information about all three factors: the opportunities that children were given at middle school to learn about science, the skills that they showed in science learning at middle school, and their SES. Structural equation models provided clear evidence that the children from higher SES backgrounds were more likely to be enrolled in Biology and Chemistry courses than other children; those with higher previous attainment in science were more able to take advantage of the opportunity of attending these courses, and consequently to make more progress in science during their time in secondary school. In summary, this longitudinal study showed that more opportunities to learn science were offered to children before secondary school, and this offer was taken up by more children from higher SES; children who had made use of the opportunity to learn science before secondary school had higher attainment in science later, when they were in secondary school.

The PISA 2015 report defines differences in "opportunity to learn" between advantaged and disadvantaged schools largely by the number of hours spent in regular science lessons and other scientific activities in the school, which account for additional variance in science scores even after taking into account individual and school SES differences. Across OECD countries, students who are not required to attend science lessons score 25 points lower in science than students who are required to attend at least one science lesson per week. Other science related school activities also had a significant effect: across OECD countries, schools that offer science competitions scored 36 points higher and those that offered science clubs scored 21 point higher than those that did not offer these opportunities. Both the requirement to attend at least one science lessons per week and the offer of science

competitions and clubs were more likely to happen in high SES schools, but these effects were significant even after accounting for the socio-economic profile of students and schools, although. "All the correlational evidence in this volume suggests that learning science at school may be more effective than learning science outside or after school. Students who spend more time learning science at school score higher in science, while this is not necessarily the case with students who spend more time learning science after school" (OECD 2016 c, pp 227). This report thus documents an opportunity gap related to the time devoted to learning science in school, which is an important factor for performing well in PISA and is associated with school SES.

Summary of the evidence on the opportunity gap at country and school level

An opportunity gap for learning science exists when investments in some students science learning is higher than investments in other students' learning. The evidence that there is an opportunity gap for learning science at the country level is robust, even though it is only based on correlational studies: richer countries, which typically invest more in education, have pupils who attain more in science towards the end of secondary school after spending the same amount of time in school as students from poorer countries, which typically invest less in education. However, this is not a simple linear trend: up to a certain level of investment, resources and attainment are correlated, but after a certain level of investment - estimated in PISA 2015 as a cumulative investment per pupil of USD 50 000 up to the end of secondary school - investment and science attainment are no longer linearly related. In countries where the resources are more equitably distributed across schools, there is less difference between schools in science attainment; in countries where the distribution of resources to schools is related to school SES, there are larger differences between schools in science attainment. The evidence for a school SES effect on science attainment in PISA is robust. This effect may be mediated by differences in resources available to teaching science in school - such as laboratories, teacher qualification, offer of science courses before secondary school, and time dedicated to teaching science in school. The correlation between school SES and resources makes it difficult to disentangle the effects of these two measures. This correlation illustrates the opportunity gap at the school level. Correlational analyses only provide weak evidence for causal relations but this sort of evidence may have been used by policy makers when new educational policies for teaching science were developed in the last two decades (see Appendix 1.4).

c. The opportunity gap at the family level

The idea of opportunity gap can also be applied at the family level, when SES is measured by considering the cultural and educational resources in the home. This was the definition used by Yang (2003), who investigated the impact of home resources on science and mathematics attainment (aggregated) in TIMSS in 17 countries (UK not included). SES was measured by a questionnaire that participants answered about ownership of items such as a musical instrument, books, encyclopaedia, own dictionary and own calculator. A total of 123,031 students and 3148 schools participated. The cultural capital factor, as he called this measure of SES, was strongly related to attainment in 15 of the countries, but not in Greece, nor in Hong Kong. The highest coefficient in the structural equation model (0.85) observed was in Slovenia; Canada, Sweden, the USA, Norway and Iceland also had fairly high coefficients, ranging from 0.49 to 0.71.

Measuring SES is always based on family indicators, but it is not always based on resources. Often SES measures are based on a factor that combines different indicators, such as family wealth with parents' education or professional status. Among the studies reviewed in this report, only two have disaggregated these factors. Areepattamannil and Kaur's (2013) study of Canadian students' performance in PISA 2006 isolated family wealth from a measure of economic, social and cultural factors. Surprisingly, family wealth had a negative and significant impact on students' science scores, but the measure that included social and cultural factors had a significant and positive effect on students' science scores. Yang's (2003) study mentioned in the previous section also disaggregated the family SES measure in two factors: a general wealth factor and a cultural factor, which is embedded in the general wealth factor but specifically refers to ownership of educational and cultural resources. This two-factor structure was confirmed in the analysis for 15 of the 17 countries (in two countries, a third factor produced a better fit) at the family SES level; at the school level, it was confirmed in only six countries and was thus not used at school level in Yang's analyses. The family level cultural coefficient was a significant predictor of the maths/science attainment scores in 14 of the 17 countries, but the results for general wealth varied widely between countries, ranging from -0.19 in Iceland to 0.45 in Hungary. Thus, a general wealth factor is less consistently related to science attainment than a specific factor related to cultural resources. We interpret this to mean that the amount of resources in a family matters, but so does the way in which the resources are spent.[11]

Summary of findings about the opportunity gap at different institutional levels

Differences in science attainment are related to resources for learning at different levels of analysis. At the country level, pupils from wealthier countries perform significantly better than those from poorer countries. At the school level, schools that invest more in science resources have pupils that perform better than those schools that invest less. At the family level, SES impacts on science learning, even when the school level effects have been taken into account. In general, there are associations between these institutional levels: wealthier countries tend to invest more in education, but this is not a simple linear relation. Case studies comparing countries reveal that some wealthier countries invest less in education per pupil. When the investment is well distributed across school within a country, the school effects are smaller. School effects seem to be mediated by material resources, such as the availability of a lab, and also by teacher characteristics and time devoted to teaching science in the school. School SES effects are significant even after taking into account family SES, and thus suggest an opportunity gap at the school level. At the family level, as far as it can be ascertained, the total wealth is less important than the social and cultural aspects of the school environment. The latter constitute the opportunity gap at the family level.

---

[11] White (1982) carried out a meta-analysis of the relation between SES and academic attainment, which does not identify science attainment separately. He reports a variation of correlations between SES measures and academic attainment between .185 and .577. In his meta-analysis, income only has a higher correlation with academic attainment (.315) than parental education only (.195); parental education and occupation have a higher correlation with educational attainment (.325) than a combination of these two factors with income (.318). These results are computed over different studies and the measures were not necessarily obtained for the same sample.

The conclusions in this section are robust; they are based on large scale studies with thoroughly validated measures of science attainment and rigorous statistical analyses. There are no notable differences in quality between the studies, even if some include a large number of countries and others focus on a few or just a single country. When studies drew unwarranted conclusions, these were not reported in the review.

SES and Motivation to Learn Science: Is There an Interest Gap?

Motivation and interest are considered key factors for attainment in educational settings. It has often been assumed that the impact of SES on educational attainment is mediated by motivation and interest, among other factors: "children growing up in families experiencing socio-economic adversity are at a greater risk to show educational underachievement and lack of motivation than their more privileged peers" (Schoon, Ross, & Martin, 2007, p. 131). DeBacker and Nelson (2000) argue that motivation and interest are particularly crucial in science. Learning many current scientific concepts involves conceptual change - i.e. abandoning one's intuitive ideas and adopting new ways of thinking that are in line with accepted wisdom in science. Such changes are "unlikely to occur if students fail to engage new information at a sufficiently deep level to recognize conflicts between existing understanding and new information" (DeBacker & Nelson, 2000, p. 245). Motivation theories distinguish between three ways of valuing something: intrinsic value, utility value, and attainment value. "Intrinsic value is a measure of one's personal enjoyment or satisfaction from engaging in tasks in the science domain. Utility value is the degree to which students value science for its usefulness in a future endeavour. Attainment value is the importance one places on accomplishments in the science domain" (DeBacker & Nelson, 2000, p. 247).

Our search of the literature did not identify many papers that included measures of SES, interest specifically in science and science attainment. Therefore, in order to ascertain the importance of considering specific interest in learning science and its correlation with SES in this review, we first examined two meta-analysis of how these two measures are related. The first, by Willson (1983), considered studies that covered a wide age range - primary school children to college students - and a wide range of publication dates - 1935 to 1980. Neither the age of participants nor the year of publication affected the magnitude of the correlation between interest and attainment. Willson provided a summary of the values of the correlations between interest and attainment: "In examining causal relationships between attitude and attainment, 42 coefficients were based on attitude measured prior to attainment ($r = 0.16$), 24 coefficients for which attainment was measured prior to attitude ($r = 0.16$), and 193 coefficients for which the two variables were measured simultaneously" (Willson, 1983, p. 844). The overall correlation between interest and attainment measures was 0.16, which is rather low. These results do not support a strong relation between specific interest in science and science attainment nor do they shed any light on the direction of causality.

A later meta-analysis by Weinburgh (1995), which combined data from 18 studies between 1970 and 1991, also investigated the correlation between specific interest in science and science attainment. It reports a correlation of 0.50 for boys (n = 561) and 0.55 for girls (n = 623). This is moderated by the type of science, with lower correlations for physics (0.35 for boys and 0.37 for girls) and higher for biology (0.53 for boys and 0.57 for girls).

Measures of interest in science have become more sophisticated over time and tend to separate out attitudes and behaviour, which were not distinguished in the previous meta-analyses. Two studies are cited here to exemplify the more recent approaches. DeBacker and Nelson (2000) studied 242 high school students who were enrolled in secondary school science classes at different levels. They measured three types of perceived value of science - intrinsic, utility, and attainment value - by means of questionnaires. Science attainment was defined by the level of science class in which the students were enrolled and by a measure of science attainment. Students enrolled in higher level courses had higher scores on the three types of value, and so did students with higher attainment scores; there were no significant interactions between these variables. The effect size for the relationship between intrinsic, utility and attainment value and science attainment was statistically significant but modest (η^2 =12). Thus, valuing science and science attainment are associated, but it should be recalled that, as the study is concurrent, it is not possible to know whether the students value and enjoy science more because they perform better or vice-versa.

PISA 2006 focused on science and included measures of attainment and also of specific interest in science. There was a great effort in measuring interest in science. The assessment included interest in different types of science related questions: (a) health related issues that student might encounter personally, such as learning that contaminated water can cause diseases; (b) interest in more abstract scientific explanations, such as learning about how the molecular structures of various plastics differ; and (c) interest in how scientific experiments are conducted, such as learning about the design of experiments to test the effects of fertilisers. These general measures of interest were complemented by items that measured embedded interest in science, which represent a different perspective to the measures used in the previous study. In these items, science attainment questions which were about understanding and applying knowledge in real-life situations were followed by questions about how interested the participant was in learning more about that topic. For example, after answering problems about the topic tobacco smoking, students were asked how interested they were in knowing more about the effects of smoking on the body. Olsen and Lie (2011) analysed responses to 52 embedded interest-in-science items and their relation to attainment in 57 countries. The embedded interest in science items were chosen because the authors expected these to produce more genuine responses than the general interest items. They do not detail their quantitative analyses of the relation between specific interest in science and science attainment but report that "contrary to what could be expected, the embedded interest scale does not correlate significantly with the students' cognitive science scale. In most countries, this correlation is close to zero. This contrasts with the broader and general measures of students' interest derived from the student questionnaire, which all have moderate positive correlations with performance in science" (Olsen & Lie, 2011, p. 106). PISA provides data collected concurrently, which cannot clarify anything with respect to the direction of causality, but the magnitude of the correlations could help identify plausible factors for longitudinal and intervention studies. The results from this analysis do not make a strong case for interest in science being a moderator of the relation between SES and science attainment.

A large scale study (N = 2,958) by Singh, Granville, and Dika (2002) was a based on the 1988 US National Longitudinal Study. Motivation and interest in science were measured by (1) attitudinal factors, such as reporting to look forward to science classes, being bored in science classes (reverse coded: i.e. larger scores in boredom corresponded to lower scores

in motivation) and thinking that science would be useful after leaving school, and (2) behavioural indicators such as time spent on science homework and time spent watching TV during week days (reverse coded). Attainment was measured by scores in a standardised test. The measure of attitudinal factors was not a predictor of science attainment, whereas the behavioural indicators measure was a significant predictor of science attainment. It is conceivable that the behavioural indicators might not always be under the participants' control (e.g. the time spent on homework may be influence by the parents' decisions), which would explain the difference between the two in predicting science outcomes. This large scale study does not provide support for the hypothesis that specific interest in learning science is a strong longitudinal predictor of science attainment.

In the studies reviewed so far, the correlations between interest in science and attainment have been examined without considering the impact of SES on interest in science. Bybee and McRae (2011) analysed the relation between interest in science and attainment at the country level and within countries in 20 countries (the UK is not included). In general, students expressed more interest in finding out about health issues than about abstract scientific explanations and how scientific experiments are conducted. There were regional variations in interest in the health issues; for example, participants from OECD countries were more interested in finding out about how air bags work than about water contamination and diseases, whereas the opposite was true of non-OECD countries. At the country level, the total interest scores were negatively and strongly associated with attainment (r=-84; our calculation from their data). Bybee and McRae (2011) report that, within countries, there was a positive correlation between interest in science and attainment, but do not report the average or the range of correlations.

Kjærnsli and Lie (2011) also worked with PISA 2006 data and included 60 countries in their data analysis. The measures of interest that they report at country level were future science orientation and interest in having a science career. At the country level, the correlations between the test scores and the interest level were both negative: r=-0.83 for future science orientation and r=-0.53 for interest in having a science career.

Using data from project ROSE (Relevance of Science Education), Sjøberg and Schreiner (2005) also carried out a country-level analysis of the relationship between a country's index of development (which consists of the per capita income and indices of literacy and of health) and students' interest in science. The interest scale had a large number of items, which included questions about how interested students were in learning about particular contents and how interested they were in having a job in science. Some specific examples given by Sjøberg and Schreiner (2005) of the correlation between interest and attainment at the country level are: (a) I would like to become a scientist (r=-0.94); (b) I would like to get a job in technology (r=-0.9); (c) Science and technology are important for society (r=-0.78) and (d) The benefits of science are greater than the harmful effects it could have (r=-0.73). The correlation between the index of development and interest in science across all items was high and negative: r= -0.85.

Shen and Tam (2008) report comparable findings at the country level of negative correlations between attainment in science in TIMSS and self-concepts about science, measured by items such as "I enjoy science" (which can be conceived of as an interest item) and "I learn science easily". These negative correlations were replicated in three waves of data collection (1995, 1999, and 2003); the nine correlations ranged from -0.37 to -0.74, which vary from moderate to quite substantial. The within-country correlations between

these items and science attainment were positive and ranged from 0.05 to 0.39, and thus were not as substantial.

In summary, correlations between specific interest in science and science attainment vary across studies; they are positive but not high. Previous meta-analyses differ with respect to the magnitude of the relationship between interest and attainment; the highest estimates are of a moderate correlation (around 0.5) and the lowest estimates are of a rather low correlation (0.16). In the PISA and TIMSS studies just reviewed, the within-country level correlations between interest and attainment were positive but low. These results are not related to SES, but they are relevant, as there might be an SES interest gap, which could mediate the SES and attainment relationship. However, at the country level, it is not possible to speak about an interest gap: the greater the country's wealth, the lower the interest in science expressed by the participants. These negative correlations are substantial and larger than the positive correlations between interest and attainment observed within the countries. In spite of the fact that there are relatively few studies at the country level, we can consider these results robust, as they involve different countries and are replicated six times (three data sets from TIMSS, one from ROSE and two from PISA).

The relation between interest in science, family SES and school SES has been analysed in three studies based on PISA data; we found no other studies that considered interest in science as an outcome of family SES.[12] Ainley and Ainley (2011) report an analysis of the correlations between SES, attainment and interest in science measured in different ways in PISA 2006: a general measure of interest in science, an embedded measure of interest in science, a measure of enjoyment of science and a measure of the perceived value of science. Data from four countries (Colombia, the US, Estonia and Sweden) were analysed, as the authors wished to sample countries that had different cultural traditions and socioeconomic levels. The correlations between family SES and the measures of interest and enjoyment of science were all negative for Colombia; for the US, one correlation was negative and the highest positive correlation was quite low (r=0.18); for Estonia, the highest correlation was 0.14 and for Sweden, 0.22. In contrast, the correlations between SES and attainment varied between 0.31 and 0.42 across the four countries. There is no information about the partial correlation between family SES and interest in science, controlling for attainment in science, but it is quite clear that there is a larger attainment gap than an interest gap.

Also working with PISA 2006 data, and including 60 countries in their analysis (aggregated by cultural similarity), Kjærnsli and Lie (2011) noted a very small effect of family SES on interest in science. They do not report numerical data, but they include in their paper a figure that displays the cumulative amount of variance in science interest explained by a measure of educational and cultural factors, plus parents' jobs, plus one of the parents holding a science related job. Inspection of the figure shows that, for most groups of countries, the cumulative amount of variance explained in interest in science by these three factors was less than 10%. They conclude that "a surprisingly weak relationship has been found between students' *future science orientation* and their home background factors.

---

[12] Schoon, Ross, and Martin's (2007) analysis of adult occupations in science, engineering and technology (SET) could be viewed as related to interest in science but adult occupation depends on attainment as well as aspirations. In this study, family SES was not a direct predictor of employment in SET occupations, and seemed mediated by aspirations expressed at age 11 for women and 16 for men.

Neither parents' actual science-related job nor other factors related to socioeconomic status seem to play an important role in the formation of students' future science professional engagement" (Kjærnsli & Lie, 2011, p. 142, italics in the original).

McConney and Perry's (2010) study, based on the results for PISA 2006 in Australia, disaggregated the data by school and individual SES both for attainment (reported earlier on; see **Error! Reference source not found.**) and interest levels. They report a lack of ssociation between school SES and students' interest in science; at the individual level, there was a weak association between SES and interest in science. Figure 3.2 displays the lines for interest in science by quintiles for individual SES as a function of school SES, also defined by quintiles.

Figure 3. 2 Level of Interest in Science Expressed in the Questionnaire in PISA 2006 by Individual and Schools SES (data from McConney & Perry, 2010)
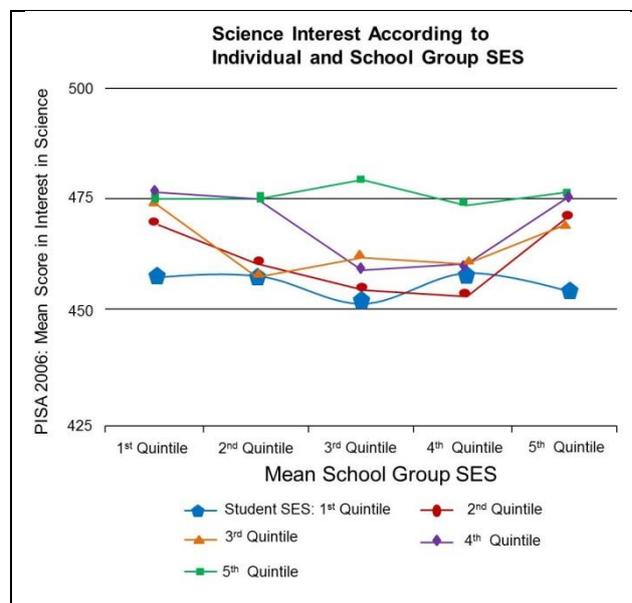
Students in the highest and in the lowest quintiles by individual SES showed the same level of interest in science irrespective of the school SES; for students in the 2nd, 3rd and 4th quintiles by individual SES, a U-shaped curve is observed. A two-level model in which student SES level was entered first and school SES second showed that student SES explained 0.5% of the variance in interest in science and school SES did not add significantly to the amount of variance explained. Unfortunately, no partial correlation controlling for attainment is reported; this would have allowed for an analysis of whether this small amount of variance was ultimately due to the effect of attainment on interest.

Summary of the evidence about interest in science as a possible mediator of the SES effects on science attainment

In summary: it has been conjectured in the literature on SES and educational attainment that motivation and interest mediate the impact of SES on educational outcomes. For this hypothesis to be supported in the domain of interest and attainment in science, two predictions should find support in the literature: (1) that interest and attainment are strongly and positively correlated (the best evidence would come from longitudinal studies where interest is measured prior to attainment); (2) that analyses of the relation between SES and interest reveal an interest gap, such that the higher the SES, the higher the interest in science. There is little evidence from longitudinal studies, so concurrent correlations were examined to analyse whether the case for the mediation of SES effects on science attainment through interest in science is plausible. The studies reviewed here provide no support for this mediation, as neither of the two predictions is upheld: the connection between interest and attainment is modest at best and the relation between SES and interest in science does not indicate an interest gap. At country level, the correlation between interest and attainment is consistently found to be strong and negative; at individual level, SES does not explain much variance in specific interest in science; at the school level, school SES does not seem to explain variance once individual level SES has been taken into account. The conclusion from these studies is that it is unlikely that raising students' interest in science will change the impact of SES on attainment.

Table 3. 1 Summary table of the evidence for possible mediators of the SES-science attainment relation found in studies that include data on SES, science attainment and the hypothetical mediator

Hypothetical mediator	Type of study	Strength of the evidence
Resources At the country level At the school level At the family level	Concurrent large scale studies using PISA and TIMSS data; concurrent large scale study in China using a national measure; one US longitudinal study of Early Childhood Education	The evidence for the impact of level of resources is very strong at the country, the school and the family levels. At the school level, SES effects are stronger when tracking is part of the educational system. Higher level of school resources can reduce the SES effect but evidence is not as robust. The effect is not inevitable: there are resilient students in all countries and some poor students in two economies (Macao and Viet Nam) outperform high SES students in OECD countries.
Interest	Large scale studies using PISA, TIMSS and ROSE data	Interest in science does not mediate the SES-science attainment relation. At the country level, the relation between interest and attainment is negative. Within countries, it is positive but modest. The relation between SES and interest in science is sometimes positive, sometimes negative, but consistently small, if significant.

Conclusions

A review of the literature on SES and science attainment led to five main conclusions.

- The impact of individual SES on pupils' science attainment is a robust effect, but it varies across countries. In PISA 2015 the amount of variance explained by individual SES among OECD countries was as low as 4.9% (Iceland) and as high as 21.4% (Hungary). The amount of variance accounted for in science attainment by individual SES in the UK was 10.5%. There effect of individual SES on science attainment is not inevitable: in the UK 35% of the students in the lowest SES level perform as well as those in the highest SES level. These results are consistent with the idea that the SES effect is mediated by a third factor, which is usually connected with SES and science attainment.

- A country's wealth, the proportion of children living in poverty, the amount spent in education per pupil, the distribution of resources to schools, and the amount of time dedicated to science learning in school matter. In the UK, the amount of variance explained by school differences is 24% (Marks, 2006). The opportunity gap seems to be partially responsible for the impact of SES on science attainment.

- In countries where the educational system includes tracking, school differences account for more variance than those that do not use tracking. Tracking and SES are correlated. In the UK, family SES accounts for a 45% reduction in the variance explained by the pupils' location in the school system (Marks, 2006), which indicates a correlation between SES and pupil location.

- Individual SES accounts for differences in science attainment even after taking into account school differences.

- Although it was hypothesised in the literature that attainment differences could be mediated by differences in motivation and interest, this review found little support for this hypothesis. At the country level, participants in the higher scoring countries show less interest in science, irrespectively of how it was measured. Within countries, there is a positive but modest relation between interest and attainment in science. The correlation between SES and interest in science varies in magnitude, from negative correlations to small but positive correlations. On the basis of correlational studies, there is no evidence for an interest gap.

Chapter 4[13]

Exploring the cause of any SES related attainment or participation gap in science: A search for possible cognitive mediators of the SES impact on pupils' attainment in science

---

Box 4 Chapter 4 Summary

This chapter gives an account of our search for variables, that existing evidence shows are possible mediators of the SES-science attainment relation.

- In order for any variable to be recognised as a plausible candidate for being a causal mediator of the effects of SES on science learning,
    a. the variable must be reliably related to SES and to pupils' success in learning science. In principle, the correlation should be a longitudinal, predictive one.
    b. teaching interventions that improve (or weaken) pupils' scores on measures of this variable must also affect pupils' science learning in the same way.
- We have identified three such variables:
    a. children's ability to reason scientifically
    b. the level of children's literacy (morphemic knowledge and reading comprehension)
    c. children's metacognitive skills.
- So far, with each of these possible mediators, the evidence that the variable satisfies the two requirements comes from different studies, done by different research teams, using different science tasks and tests. There is a clear need to combine research on the two requirements within the same project using common science tasks.
- Another problem is the scarcity of longitudinal, predictive evidence about science learning in general and SES differences in science learning in particular. Most correlational studies of possible mediators and science attainment have been concurrent ones.

The ample evidence that pupils from higher level SES families make better progress in learning science than pupils from lower level SES families naturally prompts the question whether there are any identifiable factors in children's and adolescents' environment and experiences at home and at school that cause or exacerbate these differences. To put the question more technically: are there any possible variables that mediate the evident effects of SES differences in science learning? It is a question worth asking because we (researchers and teachers) cannot do much to alter the SES background of school pupils, but we may be able to work out how to deal with factors that hinder disadvantaged pupils' science learning.

In the last chapter we presented evidence for and against the hypothesis that variations between children from different SES levels in their interest in science might be at least partly responsible for the SES differences in science attainment and therefore might be a powerful mediator of these SES differences in attainment.

---

[13] This chapter was prepared by Peter Bryant, Terezinha Nunes, & Rossana Barros

In this chapter, we have chosen to write about three *cognitive* variables which in our view are plausible candidates for being mediator variables of this sort. They are (1) the pupils' ability to reason scientifically, (2) their literacy levels, and (3) their metacognitive skills. None of these variables has definitely been shown to mediate the relation between SES and scientific attainment at school, but there is enough evidence in each case to suggest that they might do so.

 This evidence takes two forms. The first comes from *correlational* studies in which the correlations are either concurrent (the measures that are related to each other are given at the same time) or longitudinal (predictor variables measured at one time are related to an outcome measure given at a later time). The evidence from correlational studies is stronger when it is longitudinal and also when there are controls for extraneous variables that could account for the relation between the two measures - in this case, for the link between SES and science attainment. It is commonplace to say that correlation is not the same as causation: two measures could be correlated because they are both caused by a third one. The literature about the causes of SES differences and educational attainment in contemporary societies includes the view that both social class and educational attainment are determined by a third variable, which is intellectual ability (for a discussion, see Themelis, 2008). In this case, measured intelligence would explain the link between SES and science attainment; in statistical terms, would be the mediator between SES and science attainment. This possibility is examined in Chapter 5, because the data from the longitudinal study of children in the county of Avon contains information on SES as well as measures of intelligence and science attainment.

If a variable really is a mediator of scientific learning and is part of the reason for SES differences in science learning, then it should be correlated both to SES and also to pupils' progress in science, but these correlation are not sufficient to establish that it is a mediator. In fact, there is another crucial correlational requirement for a variable to be considered as a mediator, which we shall consider, not in this chapter, but in the next (Chapter 5). If the variable is the reason why children from different SES levels vary so much in their science learning, then controlling for this variable in a regression analysis should drastically reduce the link between SES and the pupils' science scores in that analysis (see Appendix 1.5 for a definition of mediators and moderators). By controlling for this possible intervening variable you are effectively asking what the SES-science learning relation would be if the children all scored the same on a measure of this variable. If the variable is a genuine mediator of the SES- science learning connection, then that connection should not be a strong one, when the children all have the same scores in measures of the variable. We shall see in Chapter 5 how this pattern of results does occur in analyses of potential mediators.

The second kind of evidence that we will review will be *intervention* studies in which pupils' attainment in science is an outcome measure and the purpose of the intervention itself is to improve pupils' strength in the variable which might be a mediator. So if, for example, pupils' literacy is a mediator of science learning, then increasing pupils' literacy should also improve their scientific attainments. If, as a mediator, literacy accounts for some of the SES differences in pupils' science learning, the intervention should improve the science attainments by pupils from low SES backgrounds at least as much as those by pupils from high SES environments.

We have made these two kinds of evidence the basic framework for our review of potential mediators of SES differences in science not just because both are relevant measures of

causes of differences in children's learning and development, but also because we think that the only satisfactory evidence for such causal pathways is a *combination* of correlational (preferably longitudinal) studies and intervention studies. The two methods have different strengths and the strengths of one cancel out the weaknesses of the other. The strength of a longitudinal, predictive study is that it establishes a definite and measurable connection between two or more variables, but the weakness is that it does not establish that the connection is a causal one. Intervention studies have the opposite strength and the opposite weakness. If the study is well designed and well controlled, a successful intervention (i.e. the intervention group scoring higher than the control group after the intervention) demonstrates a causal connection within the intervention, but its weakness is that there is no guarantee that this connection plays any part in real life. Together the two kinds of study can provide formidable evidence for, or against, the existence of a causal mediator.

Projects that combine longitudinal correlations and interventions are, unfortunately, still very rare. They have been done in other areas of developmental psychology, namely literacy and mathematics, (Bradley & Bryant, 1983; Nunes, Bryant, Evans, Gardner, Gardner & Carraher, 2007), but, so far as we know, never in research on SES differences in science learning. However, it does happen sometimes, and satisfactorily often, that one research team studies a possible causal factor longitudinally and correlationally, and another team studies the same factor or much the same factor in an intervention study – a co-incidence (in the real sense of the word) that makes it possible to draw convincing positive or negative conclusions about causal mediators. In our review of possible mediators between SES and science attainment we have searched for this kind of coincidence, and have occasionally found it.

We will describe our search for correlational and intervention studies in the following sections on possible mediators. In each section we will start by introducing a possible mediator of science learning and will first describe why people have suggested it as a possible cause of children's progress in science : then we will look at the correlational evidence for the mediator, and this evidence will always include correlations between the mediator and children's scientific skills and knowledge, and sometimes, though not always, between these variables and measures of the children's SES: finally we will present the evidence from intervention studies which compare progress in learning about science in a group (or groups) of children who are taught about the mediator with a control group (or groups) of children who are taught about something else.

Theories about scientific reasoning (the control of variables strategy (CVS)) as a mediator

Research on children's scientific understanding began with the idea that pupils' ability to reason about the design of scientific experiments might play a powerful role in children's learning about science. Inhelder and Piaget (1958) proposed that scientific reasoning depended on a sophisticated form of thinking called "formal operations" which, they claimed, is acquired in late adolescence. Formal operations involve thinking explicitly about one's own thought processes, and Inhelder and Piaget argued that the ability to do so is an essential requirement for planning scientific experiments and for understanding how to judge the value of experimental results. In their own research they concentrated on children's ability to plan well controlled tests of cause and effect, and they devised a set of tasks in which children had to work out what variables determined particular factors.

We have to make distinction here between scientific reasoning and science learning in general. Scientific reasoning is about how pupils reason about and evaluate the evidence that they are given, and in particular how they isolate and identify scientific variables. It is an essential part of planning well controlled experiments. Science learning is much more than that since it involves acquiring scientific knowledge, often without much information about how it is gathered in the first place.

In each of the tasks that Inhelder and Piaget devised the children were asked to work out what determines a particular physical outcome, such as:

- the speed of a pendulum's oscillations,
- whether an object floats or sinks when put in a container of liquid,
- the flexibility of different rods,
- balance on a balance scale
- the distance covered by a toy truck on a flat surface after it has rolled down a slope.

In each of these and in other tasks, Inhelder and Piaget provided children with the opportunity to test the effects of a several, different variables (for example, in the pendulum problem, they could vary the length of the string, the weight at the end of the string and the force of the push to start the swinging). The main aim of these studies was to see whether the children can manage to test one variable at a time, keeping all the other variables equal. Inhelder and Piaget worked with young people aged from 6 years to 18 years and found that some of the older ones did adopt this essential strategy, which nowadays is usually called the "control of variables strategy (CVS)", but others, including most of the younger children, did not. This developmental difference has been confirmed many times (e.g. Bullock, Sodian & Koerber, 2009), though there is some debate on whether in more favourable contexts younger children can manage to set up a controlled scientific comparison: a study by van der Graaf, Segers and Verhoeven (2015) shows that the age at which children do adopt the CVS is affected both by variations in their non-verbal intelligence and by the complexity (the number of possible variables) of the task itself.

An important aspect of Inhelder and Piaget's ingenious and influential research is that their tasks were *active* ones in the sense that the children had to construct the experiments (on the pendulum, on floating and sinking etc.) for themselves, which we (Bryant, Nunes, Hillier, Gilroy & Barros, 2015) call *Create tasks*. Quite a lot of research has since been done on a *passive* version of the CVS task, in which children are told about comparisons made by other people, some of which are well controlled and others not, and are asked to judge their value: we call these *Evaluate tasks*. Evaluate CVS tasks are still quite difficult (Chen & Klahr, 1999), particularly for young children, but Bullock, Sodian and Koerber (2009) showed that they are consistently easier than Create CVS tasks; this difference, which was confirmed by Bryant et al. (2015), raises the question of the relative validity of these two approaches to CVS. Which is the better predictor of children's abilities in science – Evaluate tasks or the Create CVS tasks?

Another essential aspect of Inhelder and Piaget's ideas about children leaning science is the emphasis that they gave to children's reactions to contradictions. This can be found in the section of their book which discusses how children predict which objects in a set provided by the researchers will sink and which will float when placed in a tub of water, and how they explain the basis of their reasoning. Most children, as well as many adults, do not approach this task with a ready-made set of classifications and explanations. They create the

classification of sinking and floating objects as they solve the task. Younger children tend to think that small objects float and large ones sink; many primary school children, adolescents and adults think that heavy objects sink and light ones float. After the participants have made their predictions and have explained why the objects in one group float and those in the other sink, they are confronted with contradictions, either in their classification or in their predictions. Inhelder and Piaget analysed their reactions to contradictions, which is central to scientific reasoning: if a scientist makes a prediction and it turns out to be wrong, how does this affect the scientist's theoretical assumptions? The ability to deal with contradictions, the ability to control variables in experiments and the ability to make inferences about associations between variables from data are three examples of scientific reasoning that are reasonably expected to affect how learners discard their intuitive but incorrect conceptions and take on board the accepted wisdom in science as a coherent view which is consistent with observations.

a. Correlational evidence on scientific reasoning

Our search for evidence about the relationships between children's scores in scientific reasoning tasks on one hand and their scientific attainment on the other hand has led us to two studies only. The first is longitudinal evidence reported in 1999 by Bullock & Ziegler on scientific reasoning (as part of the remarkable Max Planck project which followed the same cohort of pupils in schools in Munich over 20 years). Bullock & Ziegler found a steady improvement from the ages of 9 to 23 years in the participants' scores in CVS tasks, which provides a longitudinal validation for the developmental path suggested by Inhelder and Piaget. They also showed that these CVS scores predicted the pupils' attainment in a measure of scientific knowledge and of the participants' understanding of the way that scientists work. These were measures designed by the researchers, rather than a measure of science attainment in school, which is a great pity because the data to connect the research on the pupils' CVS scores with their actual progress in learning science at school were probably available.

The second relevant correlational study was carried out by our own team. In a large-scale longitudinal study, Bryant et al. (2015) found that the scores of over a thousand children in an Evaluate and also in a Create CVS task given to them at the ages of 10 – 11 years predicted their performance significantly in the nationally standardised Key Stage 2 Science test, given to them at roughly the same time as the CVS task, and also in the Key Stage 3 Science test, which they took three years later on (at the age of 14 years), even after stringent controls for differences in intelligence. In these regressions the predictive power of the Create scores was greater than that of the Evaluate scores. These results support the idea that pupils' ability to reason scientifically plays an important and lasting role in the progress that they make in learning about science at school.

The results of several other studies have supported the conclusion that the strength of pupils' understanding of the nature of well-controlled experiments is related to their success in science learning (Lawson, 1983; Bitner, 1991; Shayer, 1999) but none of these contained any control for the effects of differences in intelligence.

Thus, correlational research into the CVS measure does suggest that children's ability to reason about scientific experiments affects their science learning. This raises the question whether there are also differences between children of different SES levels.

b. Scientific Reasoning and SES

Here the evidence is sparse but fairly consistent. Very few studies have provided any direct information on differences or similarities between the scientific reasoning skills of children from different socioeconomic backgrounds, but all the data point in the same direction. Children from more prosperous backgrounds do better than those from poorer backgrounds in a variety of tests of scientific reasoning. Some time ago, Karplus, Karplus, Formisano & Paulsen (1977) carried out a quite ingenious forerunner of current international comparisons of children's scientific abilities by comparing children from seven different countries in a CVS task and also a task in which they had to calculate proportions. The researchers estimated the socioeconomic status of the children on the basis of the school that they were in and the organization of schools in each country. In the UK, for example, the comparison was between children in Direct Grant, Grammar and Comprehensive schools. The children in the Direct Grant schools had the highest scores in the CVS task, the children in the Comprehensive schools the lowest scores while the Grammar School pupils' scores were in between the two. This result gives us no more than a very tentative suggestion of socioeconomic differences in understanding the need to control variables partly because, ironically enough, the researchers in what can only be described as a lapse in CVS made no attempt to control for possible differences in other variables - in the children's measured intelligence, for example - and partly because there must have been a huge overlap in SES between the children in the different kinds of schools.

Finally, we return to our own large scale longitudinal study (Bryant et al, 2015) mentioned above which showed that children's CVS scores predicted their science attainment over the next three years. Further analysis of the scores in the same study has shown that the pupils' SES was significantly related to their scores in the CVS task even after controls for the effects of differences in intelligence. This seems to be the strongest existing evidence for a relation between pupils' SES and their scientific reasoning.

c. Intervention studies

In this and in other sections on intervention we will ask two questions. One is whether it is possible to improve, through a teaching intervention, the skill represented by the mediator – in this case to improve children's performance in CVS tasks. The second question is whether improvement in the mediating skill prompts any improvement in children's science attainment at school.

Several studies provide evidence that it is possible to improve children's use of CVS through direct instruction (Chen & Klahr, 1999; Klahr & Nigam, 2004; Adey & Share, 1990; Oliver, Venville & Adey, 2012; Cattle & Howie, 2008) and repeated experience of solving CVS problems (Siegler & Liebert, 1974). The most striking of these is the intervention study by Klahr and Nigam who taught two groups of children about investigating the factors that determined how far a ball that had rolled down an inclined place would continue to roll after it had reached the bottom of the plane. The possible factors were the length of the inclined plane and its slope, and (in a post-test) the roughness of the surface that the ball rolls on after it has left the inclined plane. In a one-day intervention, one group was taught by direct instruction: this group designed tests, which the investigator then discussed with them, explaining which were well controlled tests and which were not. The other group was taught by what was misleadingly called the "discovery" method: the children in this group also designed a series of tests but the investigator did not point out the children's successes

or failures to them. (Discovery teaching methods, it should be noted, do not exclude feedback from the teacher.) Many more children in the direct instruction group than in the discovery group succeeded in designing well controlled scientific tests. This difference, however, was not at all surprising since the only difference between the two groups was that the direct instruction group was given feedback by the investigator, while those in the discovery group were not.   In a following study, Toth, Klahr and Chen (2000) took a version of this successful intervention into the classroom: the regular classroom teacher gave direct instruction about how to control variables in scientific investigations. However, this study had no control group, and we can only conclude that there is a need for a well-designed experiment along these lines.

Turning to the second question, we could find little data from scientific reasoning intervention projects about their effects on pupils' progress in science at school. One research team, headed by Shayer and Adey, did make systematic attempts to establish whether interventions that successfully improved pupils' performance in the CVS tasks also bolstered their levels of attainment in science at school, but with chequered results. In one report Adey and Shayer (1990) described the results of an intervention designed to improve children's use of CVS strategies in most of the CVS tasks devised by Inhelder and Piaget (1958). The intervention consisted of 30 training sessions spread over a period of two school years and its results showed some improvement on the part of the children given this extra training in comparison to children in a control group who did not receive this training. However, there were no systematic differences between the children given the intervention and the control group children in tests of science attainment at school.

d. Conclusions about scientific reasoning (the control of variables strategy) as a possible mediator of SES differences in learning science

- Studies of children's CVS abilities suggest that their ability to reason scientifically may have a profound effect on their progress in science at school and may very well also mediate SES differences in scientific learning (Tables 4.1 & 4.2).
- But, before we can draw any firm conclusions along these lines, there has to be a great deal more research on the correlations between SES levels and scientific reasoning and on the effects of scientific reasoning interventions on children with different SES backgrounds.

Literacy and science learning

Theories about literacy as a mediator of science learning

Traditionally in educational research, children's literacy has been linked much more often to their learning about arts subjects, such as English (in English speaking countries), than to their learning about science or mathematics. Much of the teaching of English in UK schools is about the language itself and so it makes sense to consider that a child's ability to read and write the language will play a crucial part in his or her progress in the subject. There is even some empirical justification for taking the view that literacy is more important in learning about English than about science. A longitudinal study of data in the ALSPAC project (Bryant, Nunes & Barros, 2014) has shown that 8- and 9-year-old children's knowledge of phonologically based spelling rules and their understanding of the role of morphology in spelling were much better predictors of the children's performance in Key Stage English than in Key Stage Science tests over the next 5 years (KS2 and KS3).

However, that particular study did also show a consistent positive relationship between children's scores in these two basic reading and spelling tasks and their level of attainment in science over a five year period, and in recent years a great deal of evidence has accumulated of a positive connection between children's reading and writing and the progress that they make in learning about science. This connection is an important one, but, as we shall see, the reasons for it are not yet completely clear.

There are several suggestions why a pupils' reading ability might have an important effect on how well they do in learning science. We shall concentrate on two of these. One, summarised well by Norris and Philips (2003), is that scientific knowledge today is entirely dependent on the existence of scientific text, and therefore that pupils need to understand and eventually to write such texts themselves. These authors, and others (e.g. Cromley, 2009), also argue that reading comprehension and science understanding are closely related because both depend on creating and analysing coherent arguments and on making inferences

The second is that scientific texts pose particular problems to learners because they contain unfamiliar words and complex grammatical constructions. These possible obstacles have been described quite comprehensively in discussion articles by Fang (2006) and by Snow (2010) both of whom are concerned about the need in scientific texts to use precise technical terms such as *deciduous* and *torque* that are likely to be new to pupils reading scientific texts for the first time.

These terms are often multi-morphemic and it takes some understanding of the morphemic structure of words to work out how different scientific words are related to each other. This is illustrated by an example in Fang's article: the words *compress*, *compression*, and *compressional* all have the same root but the- *ion* ending in *compression* turns the verb *compress* into an abstract noun, and the *–al* ending in *compressional* makes it an adjective. The sight of these words in written form is probably a powerful way to learn about their common root (in this case *compress*) and also how affix morphemes, like *–ion* and *–al*, affect the grammatical status of the words children are reading. It is only fairly recently that literacy experts have shown that learning about the morphemic structure of multi-morphemic words and about morphemic spelling rules is an important part of children's reading comprehension (Nunes, Bryant & Barros, 2012). Fang's claim is that this kind of morphemic knowledge is particularly necessary in science learning.

a. Correlational evidence on literacy and science

The correlational evidence of a link between children's literacy skills and their success or lack of it in learning science is abundant, but most of the studies that demonstrate this link suffer from a particular weakness. In most of this research, the two kinds of attainment, reading ability and science learning, were measured at one and the same time. We have already discussed the limitations of this kind of concurrent correlational research in Chapter 3.

In 2009 Cromley published a landmark report based on three different PISA rounds (PISA: 2000, 2003 & 2006). She correlated the participants' reading scores and their science attainment scores in each of the countries taking part in the round (44 countries in PISA 2000: 41 in PISA 2003: 57 in 2006). Note that this was a concurrent correlational study: all the measures were given at roughly the same time. The overall correlation between reading

and science scores was extremely high, which certainly suggests a link between learning about reading and about science, but the strength of the correlation varied a great deal between countries. Cromley noted that the reading-science correlation tended to be higher in countries with relatively high average reading scores than in those with lower average scores.

She ran regressions on the relationship between the countries' average reading levels and the overall correlations within each country between children's reading levels and their attainment in science. These analyses strongly supported her observation of a highly significant connection between the countries' overall reading levels and the connections that exists between school children's reading levels and their success in learning about science. This last conclusion has to be a tentative one since it rests on the assumption that the reading test results are really comparable across countries despite differences in languages and in orthographies, but the study does establish first that there is an important link between how well children read and how successful they are in science at school, and second that this connection can vary quite a lot across different contexts.

A great deal of research supports the first of these conclusions (the reading-science connection), and some further research supports the second (much variability in the strength of this connection). Another large-scale concurrent correlational study of the relationship between pupils' reading and their understanding of science was carried out in the USA by Maerton-Rivera, Myers, Lee & Penfield (2010). Their study was of 23,854 5th graders (10 year-olds) in a total of 198 schools. All of these children were given school tests of reading and of science, and the researchers found that there was a strong and significant relationship between the children's reading ability and their science scores (the coefficient for this relationship in a multi-level model analysis was .51). There was also an interaction involving whether the children were learning English as a second language or not (co-efficient of .08 in the same analysis) which was due to fact that the relation between children's reading ability and their science scores was slightly weaker for the children who were learning English as a second language than for children who were not. To use Baron and Kenny's (1986) terminology, having English as a second language is a moderator of this relation.

 Several concurrent correlational studies confirm the link between reading and their scientific knowledge.   In a large scale concurrent study, Cano, Garcia, Berben, and Justicia (2014) measured reading comprehension in 604 14 year-old Spanish children and their science attainment, and several other variables which included measures of the children's "approaches to learning". They found a strong, direct relationship between pupils' reading comprehension and their attainment in science at school, and they also found that aspects of their learning approaches (such as whether they adopted a deep or a surface approach to learning) were related both to the pupils' reading comprehension scores and to their science attainment.

A similar result was reported by Van Laere, Aesaert, and van Braak (2014) whose concurrent correlational study involved 1,761 Belgian 9-year-old pupils. The researchers gave them a reading comprehension test, and they report a strong and significant correlation between the pupils' scores in this test and in a science test.

In the Netherlands, Korpershoek, Kuyper and van der Werf (2015) measured the reading ability of a large group of pupils who were near the end of their secondary school career,

which can be five or six years, and had completed courses in Physics and Chemistry. The reading ability measure took the form of a wide ranging set of tests that included a sub-test of verbal intelligence as well as reading comprehension and writing tasks. Some of the participating pupils were taking courses designed to prepare them for university (Track A) while others were being prepared for professional courses (Track B,) and the actual science courses that they had just completed were different (the courses taken by Track A were more difficult than those taken by Track B). The researchers found a strong relationship between the Track A pupils' reading ability and their performance in the end-of-course exams in Physics and Chemistry, but they did not observe the same result with the Track B students, whose reading scores and attainment levels in science were largely independent of each other. This difference echoes the variability in the reading-science relation that was found in the Cromley studies. There is a striking link between how well children read and how well they do in science classes at school, but it is sensitive to context in ways that still need to be explored.

So far in this section we have dealt with concurrent studies only. These studies are useful, in that they have established that two forms of learning, which for a long time were thought of as quite disparate, are in some way connected, but the nature of the evidence in these concurrent studies makes it impossible for us to go beyond this limited conclusion, because we cannot be sure of the direction of cause and effect in such studies. We need longitudinal studies and intervention studies to provide that basis. In Chapter 5 we will discuss the results of longitudinal research of our own that goes some way to doing that.

b. Literacy and SES

In the theoretical section on literacy we concentrated on two aspects of literacy development as possible mediators of science learning: (1) understanding the morphemic structure of multi-morphemic words and of the effects of morphemes on spelling and (2) reading comprehension. There is specific evidence of SES differences in both aspects of children's literacy. One of the studies that we will present as evidence for these SES effects includes literacy and science; the others are about literacy only and not directly about children learning science, so we will describe them rather briefly.

In the single longitudinal study which we found that included simultaneously SES, reading and science attainment, Morgan, Farkas, Hillemeier, and Maczuga (2016) analysed the public-use file of the Early Childhood Longitudinal Study, Kindergarten Class 1998-99, which is a nationally representative sample of kindergarten children in the US at that time. They analysed data collected at the end of kindergarten, 3rd, 5th and 8th grades. The measure of SES was a continuous, composite measure, which included father's and mother's education, occupations, and family income (described in four levels). The children answered reading, mathematics and general knowledge questions at the end of kindergarten; the general knowledge questions included topics that can be considered as related to science, such as knowledge about the earth and space, and processes in science, such as asking questions. They found a significant and consistent gap in science attainment as a function of SES already at the end of kindergarten, in the general knowledge measure. For every 1-standard deviation increase in SES, there was a .3-standard deviation increase in science attainment. The gap observed in kindergarten predicted the 3rd grade gap and then remained relatively constant from 3rd grade to 5th and 8th grade, except for the fact that the gap increased slightly for pupils in the lowest SES quintile. The gaps observed in reading

and mathematics at the end of kindergarten also predicted the science gap in third grade, but the best predictor at the end of kindergarten was general knowledge.

Morgan, Farkas, Hillemeier, and Maczuga (2016) also analysed the effect of changes over time in reading and mathematics on science attainment, and found that the amount of growth in science attainment was largely attributable to growth in reading and mathematics. The magnitude of growth in science declined by almost two thirds when growth in reading and in mathematics were taken into account; self-regulatory behavioural functioning did not have an effect on growth in science learning. It should be pointed out that the items in the reading assessment at the end of kindergarten were naturally appropriate for this stage in children's lives, such as letter knowledge, word recognition, print familiarity, identifying beginning and end sounds, receptive vocabulary and listening comprehension, whereas the assessments used in 3rd grade and later were reading comprehension measures. Thus the low predictive value of reading at the end of the kindergarten (β=.09) might be explained by the nature of the items. When the later tests were taken into account, a much more substantial value was found in the prediction of slopes in science attainment (β=.24 for reading and β=.26 for maths). The longitudinal nature of the study and its large and representative sample in the US (N>7,350 for these analyses) makes this an important result in the search for evidence that relates SES, reading comprehension and scientific learning.

An impressive project on SES differences in reading comprehension was done by Kieffer (2012) who studied 9,189 American children's reading comprehension as part of a set of measures on literacy given to the children on seven different occasions over a period of eight years: the proportion of reading comprehension tasks in the package of literacy measures used on these seven occasions was greater in the tests given when the children were older than on the first few occasions. Kieffer reported that children who came from relatively prosperous backgrounds were ahead in these assessments of children from poorer socioeconomic backgrounds at every point. He also analysed the growth of the children's reading ability over the eight years and reported that initially this was more rapid in the poorer than in the more prosperous pupils (in other words there were signs that the disadvantaged children might be catching up), but from their third year in school onwards the improvement in the children's success in reading tasks was greater among those from prosperous homes than among the poorer children.

Some of our own (as yet unpublished) analysis of data from a large-scale longitudinal study, ALSPAC, which we describe in more detail in the next chapter, confirms the existence of SES differences in reading comprehension. The relation between the pupils' SES and their reading comprehension scores was r=.343. Data from the same project established a correlation between SES and pupils scores on the test of pupils' knowledge and use of morphemic rules in reading and writing: r= .242). Thus, pupils' reading comprehension and their knowledge and use of morphemic rules in reading and writing might well be powerful mediators of the relation between pupils' socio-economic backgrounds and their level of attainment in science. Intervention studies are needed to test this hypothesis.

c. The effects of literacy interventions on science attainment at school

By and large, literacy interventions designed to improve children's science learning have been remarkably successful. However, the theoretical implications of the studies of the effects of these interventions tend to be rather limited, at any rate as far as the links

between literacy and scientific attainments are concerned. This is because in all the relevant intervention studies, the children were, for very good reasons, given comprehensive science teaching programmes which concentrated on literacy but nevertheless involved quite a bit more than just literacy teaching, and in most of the studies the researchers have not done a good job of disentangling the effects of teaching literacy from the effects of other aspects of the programme on children's science learning.

To make this point in a positive way we shall begin with the one literacy intervention study with a design that does allow us to attribute improvement in science specifically to the effects of a literacy course incorporated into the science curriculum. This was an intervention designed by Fang and Wei (2010) that lasted for 22 weeks and involved two groups of 10-11 year old children, an experimental group, who were given a literacy course as part of their science instruction, and a control group, who were taught science in the same way except that they did not take the literacy course.

Apart from this difference, the children in the two groups were taught science in the same way. All of them were taught science through an Inquiry method, which emphasised teacher-led enquiries and field-trips: the students "explored" various scientific phenomena, wrote their own reports and presented these for discussion with the other pupils.

In addition to this, once a week, the pupils in the experimental group were also given a 15-20 minute literacy class, in which they were introduced to various reading and writing strategies, such as note taking, and paraphrasing, and were encouraged to use these in the science course. They also took part in a home reading programme, which meant that they were required to take home selected science books and read and discuss them with someone else in the family. Thus, the literacy course was part of the experimental group's Inquiry based science instruction, and the only difference between the two groups was that the experimental group took part in the literacy course during the time of the project. The children in the project came from six classes, three of which were assigned to the experimental group and three to the control group.

The project was a successful one. The children's science attainment levels were measured both before and after the 22-week intervention in standardised tests, and the analysis of their scores in these evaluations showed that the children in the experimental group made more progress in science than those in the control group.

The researchers conclude from these results that the instruction on literacy, given to the experimental group, helped them to learn about science and justify this conclusion on the basis of the two groups being taught science in the same Inquiry based way, and the only difference between them being that a literacy course that was given to the children in the experimental group only. However, the children in the experimental group also took part in a home science reading programme with the help of their parents while the control group children did not, and this could have given the experimental group an extra motivational spur. One cannot be certain whether the crucial component that led to the experimental group's relative success was extra literacy or extra motivation.

The comparison between the experimental and the control groups has also been quite unclear in several other studies. One is an intervention study project by Romance and Vitale (2016): they have developed a science teaching programme called IDEAS (mercifully, not an acronym), which emphasises the enhancement of children's reading comprehension and of their writing skills, but contains many other aspects of teaching science, such as hands-on

science, reading comprehension and propositional concept mapping. In this study, a large number of children, who were taught the IDEAS program, formed the experimental, intervention group, and an equally large number was allocated to a control group and was taught science in some other way (or ways). The report sets out the IDEAS programme in some detail, and says very little about the science teaching given to the control group children in the study, but it is almost certainly the case that the emphasis on literacy was only one of many differences in the way that the two groups were taught. The researchers report that the intervention group did much better than the control group in standardised post-tests on science and also on reading. So, we can note that the IDEAS programme was a success but that there is no way of knowing how much of that was due to its emphasis on literacy.

Two other recent intervention reports from a Texas research team (Lara-Alecio, Tong, Irby, Guerrero, Huerta & Fan, 2012; Tong, Irby, Lara-Alecio, Guerrero, Fou & Huerta, 2014) leave us in much the same position of knowing that a teaching programme works but not being sure why. Both reports set out the evident success of a science teaching programme that combined Inquiry-based teaching with instruction about reading and writing. But in both cases the differences between the treatment given to the Intervention group and the Control group do not make it possible to be sure that the success was due to the combination of inquiry based instruction and the emphasis on reading, as the authors claim, since the teaching given to the control group also contained elements of Inquiry based teaching and literacy instruction.

Finally, we will mention an intervention study on the effect of combining teaching science with teaching reading, which is well designed and establishes an effect of the intervention on pupils' reading ability but leaves us in the dark about learning science. This is a much cited study by Morrow, Pressley, Smith and Smith (1997) on a programme that combined inquiry based science teaching with instruction about literacy. Six classes of 8-year old children took part in the study and the classes were randomly allocated to three groups. In one group the pupils were given the combined science and reading programme over the whole of one school year; in another they were given just the literacy part of the programme; the third group was a control group not given any other teaching than the normal school courses. The outcomes of this study were mostly positive. On the literacy measures the children given the combined science and literacy programme did better in the post-tests than those who experienced just the literacy part of this programme. The children taking the combined course also did better on two out of three science post-test measures than those in the control group. These results clearly establish that combining literacy with science teaching was a better way of teaching literacy than just teaching them literacy, and that certainly is an interesting outcome. But it is about learning literacy, and not about learning science. This is because the design of the study was asymmetrical: it did not include a science-only group of children. So, we cannot tell whether the group given the combined programme had higher scores in the science post-tests because the programme combined science with literacy teaching or just because it included teaching on science.

d. Conclusions about literacy as a possible mediator of SES differences in learning science

- Correlational research has consistently established a strong relation between children's reading and writing abilities and their success in learning science at school (Table 4.1).

- There is some evidence that children's understanding of the morphemic structure of multi-morphemic words and also their strengths in analysing and understanding written text (reading comprehension) play an important part in this relation.

- Most of the correlational evidence comes from concurrent, rather than longitudinal research, and the people doing the research hardly ever control for the effects of extraneous variables, such as differences in the children's measured intelligence. This makes it difficult to conclude anything about the direction of cause and effect in the relation between literacy and science learning.

- Studies of interventions (Table 4.2) that combined teaching literacy with teaching science have also consistently produced positive results. However, the design of these interventions leaves some questions unanswered. It is not possible to be certain that it was the combination of the two kinds of teaching that improved the children's attainments in science in most of these studies.

Metacognition and science learning

Theories about metacognition and science learning

A common theme in theories about school education is that pupils' understanding of how they themselves think, remember and learn plays an important part in the progress that they make at school. The ability to reflect on one's own cognitive strengths and weaknesses is called "metacognition" and the initial suggestion of its significance in children's education was made by Anne Brown and her colleagues (Brown, 1980; Brown Campione & Day, 1981; Brown, Bransford, Ferrara & Campione, 1983). She argued that there are two components to metacognition: one is the knowing about cognition and the other is regulating one's own cognition. Her simple and powerful idea was that in studying, for example, human memory we should consider not just how well or badly people remember things, but also what strategies they can mobilise, such as taking notes or rehearsing and organising memories, to make it easier for them to remember what they need to know later on. She claimed that it is possible to enhance children's metacognitive skills by teaching them about these strategies.

Brown's initial research was about children reading and learning from texts, but it is easy to see that the idea could be used quite widely in education. Several authors, including Byrnes and Miller (2007), Byrnes and Miller-Coto (2016), Sperling, Howard, Miller and Murphy (2002) and Zimmermann and Kitsantas (2014) have argued that there should be a positive relation between the extent of children's metacognitive skills, particularly their self-regulatory ability, and their general progress at school. On the whole, studies of children's self-regulation (Anne Brown's second metacognitive component) in different school subjects have shown positive relations between this skill and children's educational attainments (Sperling, Howard, Miller & Murphy, 2002; Zimmermann & Kitsantas, 2014)

a. Metacognition in science: correlational data

Naturally, some educationalists have argued for the importance of metacognition in science learning – not just in reading science texts or in writing reports on scientific exercises, but

also in planning experiments and explaining data gathered in field trips and the school laboratories. Yet, correlational studies that link children's metacognitive skills with their attainment in science, which would be one good way to test these hypotheses, seem to be rather thin on the ground. The gap is partly filled by an interesting study of 941 4th to 8th grade (approximately 9 to 13 years old) Turkish children, by Topcu and Yilmaz-Tuzun (2009). These researchers related measures of the children's knowledge of cognition and regulation of cognition (Anne Brown's two components of metacognition) to their science attainment, as measured by their marks in school science tests and found strong and significant links between the two.

The success of this study is marred by the timing of the administration of the two sets of measures, which was concurrent and not longitudinal, and also by its failure to control for the impact of extraneous variables, like intelligence, that might have determined both the children's metacognitive skills and their science learning as well. These, as we have already seen, are common faults in research on science learning.

Another study from Turkey by Yerdelen-Damar and Pesman (2013) shares these two faults but presents some striking results nevertheless. It was a research project on 338 10 to 14-year old pupils who were taking a physics course at school and whom the researchers gave measures of their metacognitive skills and also of their feelings of self-efficacy in science. The researchers report that in a structural equation model, which included these three measures, the direct pathway between metacognition and the pupils' physics marks was not significant, but the indirect pathway from metacognition to self-efficacy to physics was a significant one. The researchers' explanation is certainly interesting: the greater the children's metacognitive skills, the more confident they feel about learning physics, and this confidence helps them to learn physics and get good marks in the physics exams.

This might be what was happening, but the idea of a one-way street from metacognition to self-efficacy in physics learning seems implausible to us. Because this was a concurrent study, there is nothing in it to exclude the possibility that the children's successes in physics affected their self-efficacy, or that the extent of their physics knowledge makes it easier for them to exercise their metacognitive skills in the physics course. Longitudinal versions of this study with autoregressive controls (as well as controls for extraneous variables like measured intelligence) and of the previous study would tell us about the directions of cause and effect, and in fact the main attainment of the careful and novel work that we have just described may be that it will encourage others to make the same comparisons longitudinally.

b. SES and metacognition

As far as we know, there is little research on the relation between pupils' SES and their metacognitive skills but, as it happens, both of the Turkish studies that we have just described did contain SES measures based on the pupils' parents' educational level. Topcu and Yilmaz-Tuzun (2009) reported significant relationships between the two components of the children's metacognitive abilities (knowledge and regulation of cognition) and their fathers', and also their mothers', educational levels. This strengthens the possibility of metacognition being a mediator of SES differences in children's science learning.

c. Metacognitive interventions and science learning

Over the years several research teams with an interest in metacognition have set out to look at the effects of improving children's metacognitive skills on the progress that they make in learning about science at school. Some of these, for example a study by Baird (1986), are qualitative studies without control groups, which look at changes in children's scientific knowledge at a time when they are being encouraged to use metacognitive skills appropriate to the scientific topics being covered in their science classes. These usually contain some fascinating observations of children's comments and solutions to scientific problems, but they do not fall within the remit of this report.

We will start our account of interventions with control groups by describing a study by Zohar and Peled (2008) on 10-year old Israeli pupils learning about seed germination over five sessions using a computer microworld. The 41 children in this intervention project were categorised on the basis of their school results in science as high or low achievers and they were further divided into an experimental group of pupils who were given instruction about the appropriate metacognitive strategies to use in the microworld task. This task was about the control of variables: the children could vary various factors and control others to determine what led to successful germination; the metacognitive instruction was designed to help them see that the solution to the task was to vary one variable at a time while controlling the others. The children in the control group were given the same task also in five sessions, but not the metacognitive instruction.

As predicted, the children in the experimental group adopted the correct control of variables strategy sooner and more effectively than those in the control group and applied it to two transfer tasks (one very like and the other very dissimilar to the original training task) better than the control group children did. One interesting additional result was that the high achievers were more successful than the low achievers in the initial sessions in both groups, but within the experimental group (taught about metacognitive strategies) the improvement in the low achievers' solutions was much greater than it was with the high achievers, so that by the last of the five sessions the low achievers had almost caught up with the high achievers. The result suggests that metacognitive instruction might be one good way of narrowing the gap between disadvantaged and other pupils in science lessons.

This interesting pattern of low achievers benefitting more than high achievers from metacognitive instruction has also been reported in a study by White and Frederiksen (1998) on metacognition in a physics course. The 7th to 9th grade pupils participants in this study were around 360 in number (the article is vague on this point). They were given a 10.5 week course on Newtonian physics, which was administered by their science teachers and was called the ThinkerTools Inquiry Curriculum. The course itself involved some metacognitive instruction, since it contained instruction about forming clear hypotheses, generating alternative hypotheses and applying any conceptual models that they develop to new contexts.

In addition to this, some of the pupils (about 50% of the students in each of the 12 classes taking the ThinkerTools Inquiry Curriculum) were also given several sessions on Reflective Assessment, in which they were asked to judge the worth and standards of the work that had been done in the science classes. They were taught to reflect on "goal-oriented criteria such as *Understanding the Science* and *Understanding the Processes of Inquiry*, process-

oriented criteria such as *Being Systematic* and *Reasoning Carefully* and socially-oriented criteria such as *Communicating Well* and *Teamwork"* (White & Frederiksen, 1998, p. 24).

White and Frederiksen reported a pattern of results remarkably like the pattern found in the Zohar and Peled study. Assessments given during the intervention and post-tests of knowledge gained during the course showed that pupils in the Reflective Assessment group did better than the Control group and also that, within the Reflective Assessment group, the improvements in the scores of the low achievers were significantly greater than those of the high achievers.

d. Conclusions about metacognitive skills as a possible mediator of SES differences in learning science

- Research on metacognition fulfils all our requirements for a possible mediator of SES differences in science learning. Pupils' metacognitive skills are related to how well they do in science at school and instruction in metacognition does improve how successful they are in science tasks.

- In both the intervention studies reported here, the effects of the metacognitive interventions were particularly strong with low achieving children. Since children from low SES backgrounds were almost certainly over-represented in the low achieving groups, these two studies certainly demonstrate the relevance of metacognition to the question of SES differences in science learning.

- We also presented evidence of SES differences in metacognitive skills, though we must acknowledge here that this is the weakest link in our argument because we were only able to find one study that demonstrated this relation, and this was a concurrent correlational study with all the weaknesses of that genre. There is a need for good longitudinal research on the role of metacognition in learning about science.

Other possible mediators of SES differences in learning science

In this chapter we chose the three candidates for mediatorship (to coin a new term) of the SES–science learning relation on the basis of the amount of evidence pointing in their direction. As we have seen, this evidence was not conclusive for any of the three possible mediators that we discussed, but in each of the three cases it seemed convincing enough to warrant further research.

We were not confident enough about any of the other possibilities that we looked at to recommend them as suitable candidates for immediate longitudinal research or full-scale intervention studies. With some variables, like children's interest in science or their science motivation, the evidence (reviewed in Chapter 2 of this report) pointed to there being no strong or reliable relation between the variable and SES.

Some variables look more promising but at the moment there is not enough information yet about their links with SES and with science learning to be sure that they qualify for immediate research as mediators. One of these is vocabulary. Learning science demands the understanding of a large number of new words, and some research has already established the possibility of studying children's vocabulary in relation to the scientific vocabulary that teachers use in science classes on specific topics. An example is research on teaching about physical density which establishes that there are identifiable gaps in the pupils' knowledge of the meaning of technical terms that teachers use to instruct children

(Seah, Clarke & Hart, 2011; Xu & Clarke, 2012; Seah, Clarke & Hart, 2015).   These were comparison studies, but there is also an obvious opportunity here for longitudinal, correlational research between measures of pupils' vocabulary and their success in science learning later on, which, so far as we know has not been taken, apart from some research on children's perception of the importance of vocabulary in science learning (Brown & Concannon, 2016).

In Chapter 1 we wrote about some research in the Netherlands by Uerz, Dekkers, & Béguin (2004) on participation in science courses, which of course is a different topic to science attainment, but is worth mentioning here because it throws some light on the complexity of the relation between pupils' language skills and their science learning. The authors measured the "gap" between pupils' scores in standardised assessments of their skills in language and in mathematics and reported that the extent to which children's mathematics scores outstripped their language scores (the gap) was an excellent predictor of how likely pupils were to choose to specialise in science. This relationship between the gap and science choices remained significant even after controls for the pupils' actual scores in the mathematics task. The researchers argued therefore that relatively high mathematical skills lead children to opt for science courses, but they did not follow the logical implication of their own argument for the role of language skills in science participation, which is that relatively high language skills actually deter children from enrolling in science.

Oral language nevertheless lies at the heart of several new and apparently successful ways of teaching science, such as the use of discourse (Mercer, Dawes, Wegerif & Sams, 2004; Howe, Ilie, Guardia, Hofmann, Mercer, & Riga; 2015) and argumentation between pupils (Erduran, Simon, & Osborne, 2004; Osborne, Erduran, & Simon, 2004; Osborne, 2010; Osborne, & Patterson, 2011), but the effects of these promising new methods on SES differences in science attainment are not yet known. Nor could we find any adequate longitudinal research on the impact of children's skills in argumentation and discourse on science learning or on SES differences in these skills.

Another promising topic is the well-researched idea of "conceptual change", which is based on the recognition that pupils come to science lessons initially with their own informal and often implicit concepts about scientific phenomena (Duit & Treagust, 2003). Learning science, according to those who pursue this approach, is a matter not simply of pupils acquiring new concepts and new ways of thinking about the physical and biological world, but also of them adjusting and to some extent abandoning their previous understandings (Amin, Smith & Wiser, 2014). We were unable to find research on SES differences in this field or on any predictive relationship between these early informal and incorrect understandings and science attainment later on.

Table 4. 1 Summary table of the contribution of the correlational studies towards establishing variables that might determine the relationship between SES and science attainment

	Relation to SES	Prediction of science attainment	Concurrent or longitudinal design	Control for extraneous variables
Scientific reasoning	Sparse evidence for a connection between SES & scientific reasoning	Much of the evidence that children's scientific reasoning predicts children's science learning is based on performance in science tasks devised by researchers, not on standardised tests of science attainment	We could locate one longitudinal study only of the relationship between scientific reasoning and science learning (apart from our own longitudinal ALSPAC study reported in Chapter 5 of the report)	Correlational studies of scientific reasoning and science attainment rarely control for extraneous variables like measured intelligence but the longitudinal study reviewed did so.
Literacy: reading comprehension	Consistent evidence of a relation between reading comprehension & SES	Consistent evidence of a strong relationship with science attainment. Most of this evidence in based on standardised measures of science attainment	Most of the evidence on reading comprehension as a predictor of science attainment takes the form of concurrent correlational data	Controls for the effect of measured intelligence differences are extremely rare both in studies relating SES and reading comprehension and in those relating reading comprehension and science attainment
Literacy: morphemic knowledge	There is only one (as yet unpublished) study of the relation between pupils' SES and their knowledge of morphemic spelling rules	Some evidence exists of a longitudinal relationship between children's use of morphemic spelling rules in reading and writing and science attainment	The evidence linking knowledge of morphemic spelling rules and science learning is longitudinal.	The evidence linking knowledge of morphemic spelling rules and science learning does include controls for the effects of measured intelligence differences
Metacognition	Research connecting SES and metacognition is sparse, but it does suggest a strong relation between the two variables	There is consistent correlational evidence for a relation between children's metacognitive skills and their attainments in science	All of the studies relating metacognition to science attainment that we could find are concurrent ones	None of the correlational studies that we reviewed contained controls for measured intelligence or other extraneous variables

Table 4. 2 Summary table of the contribution of the intervention studies towards establishing variables that might determine the relationship between SES and science attainment

	Design of the intervention studies	The length of the interventions	The nature of the outcome measure
Scientific reasoning	Mostly quasi - experimental with the random assignment of classes to different groups/conditions	The interventions vary in length from very short to very long (e.g. the two year intervention in the CASE study)	The outcome measures are mostly designed by the researchers and do not involve standardised measures. Some studies have used performance in national exams as well
Literacy: reading comprehension	Most of the reading comprehension intervention studies with science as the outcome measure have quasi-experimental design and most involve intervention of another variable, such as the use of inquiry methods (besides reading comprehension), and do not control for this other variable	These interventions are adequate in length	Some studies do involve standardised measures of science attainment
Metacognition	Intervention studies on metacognition have been about self-regulation: they involve quasi experimental designs with random assignment of classes to different conditions	The interventions are adequate in length	The research studies on metacognition and science show not only a positive effect of instruction on self-regulation, but also a narrowing of the gap in science learning between high and low achievers as a results of this intervention

Conclusions about possible cognitive mediators of SES differences in science learning

- Using a relatively loose set of criteria, we identified three variables (Tables 4.1 & 4.2) that seem to be plausible candidates for mediating SES differences in science attainment. These were (a) scientific reasoning, (b) literacy, and (c) metacognition.
- There are two ways, at least, in which literacy skills could mediate the effects of SES on science attainment: one is through learning about the morphemic structure of scientific terms, and the other is through reading comprehension.
- Research on all three of the possible mediators is marred by a lack of longitudinal studies between the variable in question and pupils' science learning.

Chapter 5

Exploring the cause of SES-related attainment or participation gap in science: An analysis of the ALSPAC data on science attainment in KS2 and KS3[14]

Box 5 Chapter 5 Summary

Chapter 5 presents the results of fresh analyses of information in the database from the Avon Longitudinal Study of Parents and Children (ALSPAC), which were carried out for this report in order to fill some of the gaps in existing research. The ALSPAC study was a longitudinal project with many biological, social and psychological measures on a large number of children. These variables included SES, intelligence as measured by the Wechsler Intelligence Scale for Children (WISC), reading comprehension, vocabulary, and science attainment in Key Stage 2 (KS2), a national assessment taken at 11 years, and in Key Stage 3, a national assessment taken at 14 years. The results of these new analyses are summarised below.

- Multiple regression analyses showed that both the individual SES and the school SES level predict science attainment at KS 2 and KS 3. These results indicate that the opportunity gap identified in the international literature is replicated in the UK results.
- Much of the relation between SES and pupils' attainments in science appears to be due to intervening variables. When reading comprehension was taken into account, the relation between SES and science scores was dramatically reduced; it was reduced even further if reading comprehension and vocabulary were both taken into account.
- When scientific reasoning was taken into account, the direct relation between SES and science was also reduced to a still significant but much smaller magnitude.
- When reading comprehension and scientific reasoning were entered into the equation before SES, the amount of variance explained by SES on scientific attainment went from 14.9% to 1% in KS2; in KS3, it went from 6.9% to 0.8% in Paper A (the easier paper) and from 10.9% to 2.1% in Paper B.
- These reductions in the strength of the direct relation between SES and science attainment, are good evidence that each of these variables is likely to be a mediator of the link between SES and science attainment.
- Similar analyses showed that the link of reading comprehension and scientific reasoning with science attainment cannot be accounted for by the children's intelligence.
- We conclude that the effects of SES on pupils' science attainment largely depend on reading comprehension and scientific reasoning, which is good news since the possible mediators that we have identified are highly amenable to specific and effective teaching interventions.

The dearth of longitudinal studies in which the hypotheses considered in Chapters 3 and 4 are evaluated can, to some extent, be dealt with by analysing data from the Avon Longitudinal Study of Parents and Children (ALSPAC) project. The ALSPAC data set offers an

---

[14] This chapter was prepared by Terezinha Nunes, Peter Bryant, & Rossana Barros

invaluable opportunity to surmount many of the obstacles that we have encountered in the analysis of the literature. The size of the ALSPAC sample and the longitudinal nature of the data remove the difficulties that have hindered the progress of so much previous research on differences in students' science attainment. The ALSPAC[15] study contains measures of the participants' SES, of their science attainment at two points in time, and of some of the possible mediators of the relation between SES and science attainment. This allowed for a longitudinal predictive analysis of the strength of possible mediators between SES and science learning when these variables were included in the analysis.

The measures of science attainment in the ALSPAC data are country-wide assessments known as Key Stage (KS) tests. ALSPAC includes the students' scores for two science assessments: one in KS2, which pupils take when they are about 11 years old, and the second in KS3, which they take when they are about 14 years old. The large body of data in the project also includes information about their performance in several relevant psychological tests (such as a standardised intelligence test, a standardised reading comprehension test, a standardised measure of vocabulary, and self-regulation as measured by inhibition of a response). A sub-sample of the participants was given a test of scientific reasoning in the same year in which they took KS2 tests, which was about three years before they took the KS3 tests. This measure, which is an adaptation of a Piagetian task for group administration, is known as the control of variables task. It does not assess whether the participants know anything about science concepts; it assesses their understanding that, in order to draw conclusions from an experiment, it is necessary to control for all the variables that might affect the outcome, and to vary only the variable whose influence is being tested in the experiment. The sample available for the analyses reported in this chapter varies; it includes over 5,000 participants for analyses of KS2 science attainment and over 3,000 participants for analyses of KS3 science attainment.

We first investigated the school SES effects to test whether, as in the studies reported in Chapter 3, there is evidence of an opportunity gap. We then ran a series of longitudinal analyses in which we included measures of the mediators that were considered in Chapters 4: scientific reasoning, reading comprehension, and self-regulation (which is an aspect of metacognition; the studies on metacognition described in Chapter 4 used different measures). Finally, as there is a belief amongst teachers that pupils from lower SES background struggle with science because they have poor language skills (Royal Society, 2008), we analysed whether the pupils' vocabulary could be a mediator of the SES-science attainment relation.

As a preliminary analysis, measures of each of the variables hypothesized to be mediators of the SES-science attainment link in chapter 4, was correlated with individual SES and with KS2 and KS3 science attainment. We used this initial analysis to exclude from the methods section descriptions of measures that did not meet the minimum criteria for being a mediator of the impact of SES on science attainment, which are that the measures must correlate significantly with SES and with science attainment (see Appendix 1.5 for an explanation of mediators and moderators). Because the number of participants is large, if a correlation with the KS tests or with SES is not significant at the .01 level, the measure was not included in further analyses.

Table 5.1 displays the correlations of each of the possible mediators with SES, KS2 science attainment, and KS3 science attainment. The measures of self-regulation did not meet the criteria for inclusion in further analysis, but the measures of reading comprehension, scientific reasoning and vocabulary correlated significantly with SES and with both measures of science attainment.

Table 5. 1 Correlations of possible mediators with SES, KS2 science attainment, and KS3 science attainment

	SES (maternal education)	KS2 science	KS3 science
1. Reading comprehension	.349	.541	.675
2. Scientific reasoning	.288	.446	.495
3. Self-regulation (reaction time)	-.012[a]	-.027[a]	-.032[a]
4. Self-regulation (n correct)	.019[a]	.040	.112
5. Vocabulary	.327	.522	.552
6. KS2 science	.348	1.00	.733
7. KS3 science	.417	.733	1.00

[a] p>.01

Thus this chapter examines whether the opportunity gap described in the international literature also applies in the UK and whether reading comprehension, vocabulary and scientific reasoning qualify as mediators of the impact of SES on pupils' science attainment. We now turn to the method and then to the results.

Method

The ALSPAC Sample

The ALSPAC project recruited children born between April 1991 and December 1992 in the geographical region in the West of England that was covered at the time by the Avon Health Authority. The sample is a large and representative one, since it recruited over 80% of the children born in the area during the 21-month recruitment period. Boyd, Golding, Macleod, Lawlor, Fraser, Henderson, Smith (2012) give a detailed description of the recruitment process. The sample of participants in the current study is restricted to those who participated in individual testing sessions that took place when the children were aged 8 to 9 years, in which the psychological measures were administered. The science reasoning measure was given when the children were 10-11 years, which is the same year when they took the KS2 science assessment. This measure is a concurrent predictor of KS2 science attainment and a longitudinal predictor of KS3 science attainment.

Measures

a. Socio-economic status

There are sharp socio-economic differences between families in the UK, which were reflected in the ALSPAC sample. The ALSPAC data set contains three measures that can be used in combination to assess individual SES: the mother's occupation, the father's occupation, and the highest level of education attained by the mother. The three measures are highly correlated with each other. It also contains answers to questions about educational and cultural resources owned by the children (books, computer, clock and board games). A principal component analysis showed that the measures related to parents and those related to children's own resources were described by two factors; parental factor correlated more strongly with science attainment. We compared the correlation between the factor score based on the three parental descriptors with science attainment and the correlation between the mother's highest level of educational attainment and the children's science attainment. The mother's highest level of education showed a stronger correlation with science attainment and is therefore the indicator of SES included in all the analyses reported in this chapter. The mother's highest educational level is described in five levels: CSE (Certificate of Secondary Education), Vocational, O level, A level, and Higher degree.

b. The Wechsler Intelligence Scale for Children (WISC-R)

The WISC is a well-known, and probably the most widely used, general intelligence test for children. It was administered to 7,354 children in the ALSPAC sample, whose mean age at the time was 8 years 7m (SD 3.92). The WISC assesses two aspects of intelligence, verbal (measured by six subtests) and non-verbal (measured by the remaining six subtests). However, the WISC-R manual (Wechsler, 1974) describes the measure as formed by three factors, which are called *verbal comprehension*, *perceptual organisation*, and *freedom from distractibility*. The verbal comprehension factor includes subtasks such as a measure of vocabulary, which requires defining words (e.g. What is a bicycle?), and an information task, which asks participants to provide specific information of general interest (e.g. How many weeks are in a year?). The perceptual organisation factor includes tasks such as copying a design with blocks and assembling a puzzle. The freedom from distractibility factor includes tasks such as recalling digits and implementing a coding task within a limited period of time. The administration procedures and scoring are standardised.

The analysis of language skills as a mediator of the relation between SES and scientific reasoning in this chapter will use the WISC subtask Vocabulary as a measure. The participants are asked to explain the meanings of words. Their answers are scored for accuracy on a 0-2 scale, which reflects the level of sophistication of the definition. In a factor analysis of the WISC subtests, we found that the Vocabulary score had the highest factor loading (.722) on the main factor extracted, which corresponds to verbal ability.

c. Scientific reasoning: the control of variables task (CVT)

In this Control of Variables Task (CVT - also referred to as "the inclined plane task" in the literature), participants are asked to solve the problem of what influences the distance that a wagon will travel after rolling down a ramp onto a horizontal runway. The participants do not need to know beforehand anything about inclined planes, as the instructions spell out

different possible explanation for the distance travelled by the wagon after going down the plane. In our version of the task, used in the ALSPAC project, the participants work with a set of pictures each of which is of a single wagon on a horizontal platform on the edge of a slope: the slope ends in a horizontal runway lower down. The pictures differ from each other in four different ways, each with two values:

1. The colour of the wagon (red, blue)
2. The height of the platform (one platform was twice as high as the other)
3. The load in the wagon (load, no load)
4. The texture of the lower horizontal runway (smooth, carpeted)

Teachers administered the test to their own students as a class test. The test begins with the teachers telling the students what the task is about, and in particular pointing out the four variables that they will have to think about. To start with, the teacher shows a large poster with the same 12 pictures that the pupils have on a page and tells the pupils: "The pictures show situations where the wagons will be let down the ramp. Your task is to think about the different things that could have an effect on how far the wagon goes on once it is off the ramp. Imagine that you want to test some ideas about what makes the wagon go further from the end of the ramp. Look at the pictures on the page. There are two wagons in the pictures, a red one and a blue one. We set them up either on a ramp that starts up high or on one that starts up low. Sometimes the wagons have a load, sometimes they don't. If you pay attention to the end of the ramp, you will see that sometimes the ramp finishes on a wooden surface and sometimes it finishes on a carpet".

There were two kinds of trial in the task, which we call Evaluate trials and Create trials. In the Evaluate trials, participants are asked to judge whether a comparison between two particular pictures will tell them whether one of the variables affects how far the wagon will roll. The variable is named and the child is asked, for example, whether a particular comparison will tell them whether the surface on which the wagons run at the end of the ramp affects how far the wagon rolls. In some of the Evaluate trials the correct answer is "No" because the pictures differ not just in one variable - e.g. texture-, but in some other way as well - e.g. absence and presence of a load. In others, the pictures differ only along the named dimension and the correct answer is "Yes".

In the Create questions, participants are told to look at one picture and are asked to create a well-controlled comparison by choosing the comparison picture that will establish whether a specific dimension - e.g. height of runway - determines how far the wagon will travel.

d. Reading Comprehension

The reading comprehension measure used in these analyses was the Neale Analysis of Reading Ability, Form II (NARA; Neale, 1997), which was administered to the participants when they were about 10 years (mean age 9.9 years). In this measure, the participants read a story and are asked questions about it at the end. If they misread a word, the examiner provides them with the correct word. The test is interrupted when the participants make too many reading errors, but the comprehension score is unlikely to be underestimated on the basis of their reading accuracy errors because words skipped or misread are provided. The score used in our predictions of science attainment was the number of correct answers to the reading comprehension questions.

Results

a. Individual SES and school SES: is there an opportunity gap in the UK?

In this section, we analyse the effects of the individual students' SES and of the school SES level on the students' attainment in science in KS2 and KS3. The school SES level was calculated by obtaining the average level of mothers' education in each school. For the analyses predicting KS2 science attainment, we imposed two restrictions: (1) the children had to be in the same school for at least two years, from Year 4 to 6; (2) a criterion of a minimum of 9 pupils per school was used for the school to be included in the analysis. The number of schools that remained in the analysis was 110. As pupils move to secondary school before taking the KS3 test, these criteria could not be applied to the definition of school SES in the prediction of KS3. In this case, the school attended at the time the pupils took the KS3 science test was used in the analyses.

The correlation of the pupils' science attainment in KS2 with the individual SES was .38 and with the school SES was .31; the correlation of the pupils' science attainment in KS3 with the individual SES was .36 and with the school SES it was .32; the correlation between individual SES and school SES at KS2 was .49 and at KS3 was .44. All these correlations are statistically significant and similar in magnitude; the risk of colinearity between individual and school SES is low, which indicates that it is appropriate to proceed with analyses in which the two factors are included. We focus initially on KS2 science attainment (N=2148); Table 5. 1 explains the steps in the regression analyses used in this and the subsequent sections and their rationale.

The correlations just reported indicate that individual SES accounted for 14.4% of the variance by itself and school SES accounted for 10% of the variance on its own. When included in the regression together, the two factors accounted for 16.7% of the variance (Multiple R=.409). Two hierarchical regressions indicated that each of these measures accounted for variance independently of the other: individual SES accounted for 7.4% variance after entering school SES, which shows a reduction of 7% of the individual SES effect (it is reduced by about half). This means that there is evidence for the hypothesis that school SES is one of the mediators of the impact of individual SES on science attainment. School SES accounted for 1.9% variance after entering individual SES in the equation. Thus, the amount of variance accounted for independently by individual SES is larger than that explained by school SES but both independently account for variations in pupils' science attainment.

In order to disaggregate the effects of individual and school SES, we used the same approach as McConney and Perry (2010) and separated out the individual SES into the different levels of mother's education and the school SES into quartiles because the division of school SES into quintiles revealed a very small number of pupils whose mothers had a university degree and who attended schools in the lowest SES school quintile. The analysis of variance produced a significant effect of individual SES and a significant effect of school SES; the interaction between these two terms was not significant. Thus, adopting Baron and Kenny's (1986) definition, school SES is a mediator but not a moderator of the impact of SES on pupils' science attainment (see Appendix 1.5 for an explanation of mediators and moderators). Figure 5.1 displays the distribution of scores in KS2 science per individual SES and school SES. It should be recalled that the participants were in the same school for at least two years by the time when they took the KS2 science test.

Figure 5. 1 Pupils' performance in KS2 science by School SES Quartile and Mother's highest educational qualification.
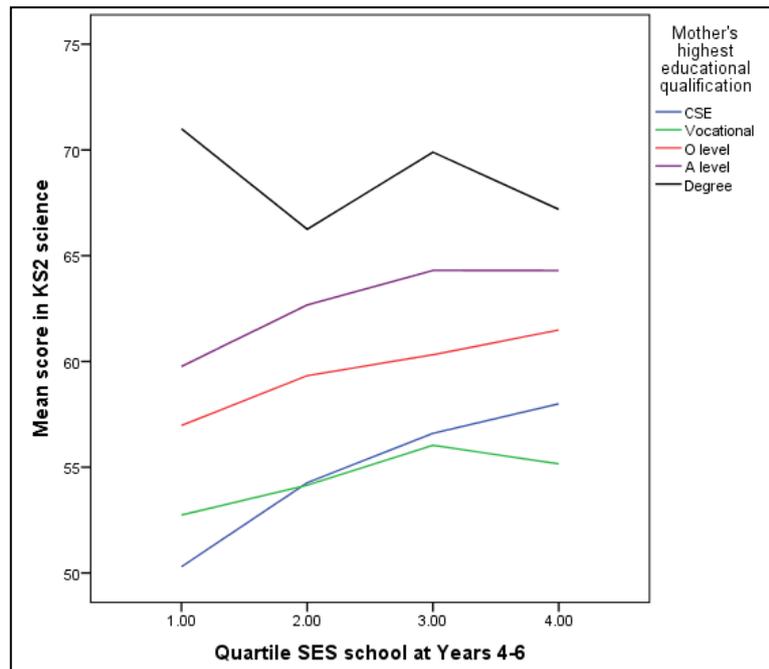
Except for students at the highest individual SES, attending a school with a higher SES in general led to better science attainment. Post-hoc tests showed that performance by pupils whose individual SES was in the 1st quintile (lowest individual SES) was significantly lower than those in the 3rd, 4th and 5th quintile, but the difference between the 1st and the 2nd quintiles was not significant. These results were echoed in the school SES effect: the difference between the 1st and the 2nd quartile (the two lowest school level SES) was not significant, but pupils in the 3rd and 4th quartile of school SES did better than those in the 1st. In brief, attending a school with higher SES decreased somewhat the gap between pupils from the lower individual SES and the highest SES in science attainment.

In summary, these analyses suggest an opportunity gap: school SES mediates to a certain degree the SES impact on science attainment. In contrast, there is no evidence that school SES is a moderator of the correlation between SES and attainment in science, as there is no significant interaction between individual and school SES.

KS3 is measured by two different examination papers, which we will designate here A and B; B is the more difficult paper. Although it is possible to run analyses with a measure that is expected to account for these differences, in previous research we (Bryant, Nunes, Hillier, Gilroy, & Barros, 2015) reported slightly different results in the analysis of how scientific reasoning predicts science attainment when the two papers are analysed separately. An exploratory analysis of the relation between SES and the science paper taken by the pupils revealed a significant association between the two (N=9,336; Chi-square = 1013.14). Figure 5.2 displays the number of pupils who took each paper by the mother's highest level of educational attainment.

Figure 5. 2 Distribution of pupils by SES who took each of the science papers in KS3.
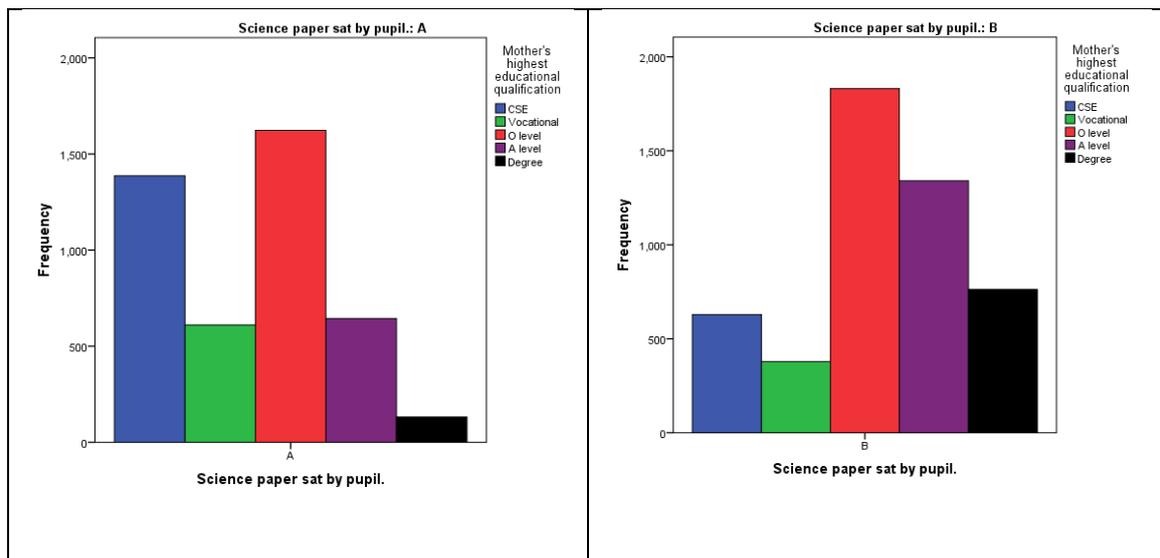
Pupils whose mothers had a higher level of education were more likely to take the more difficult than the easier paper. Thus when we carried out the analysis with KS3 science attainment to investigate school SES effects, we separated the analysis by the papers that the pupils took for KS3 science. For both papers, the analyses replicate the results for KS2 science attainment: significant effects of individual and school SES were observed and the interaction between the two SES levles was not significant.

Figure 5.3 displays the mean scores by mother's education level and school SES quartile for the easier of the two papers on the left and the more demanding paper on the right. There is a general trend for science attainment to improve with school SES, which confirms that school SES is to some extent a mediator of the SES impact on pupils' science attainment, but the individual SES effect on science attainment does not disappear. It should be noted that, for the more difficult paper, there were 576 pupils in the highest quartile for school SES but only 6 whose mothers' highest educational qualification was a vocational one; this may have distorted the mean attainment for that group, which is the lowest of all means in the figure.

Figure 5. 3 Pupils' performance in KS3 science in Paper A (left) and Paper B (right) by School SES Quartile and Mother's highest educational qualification.
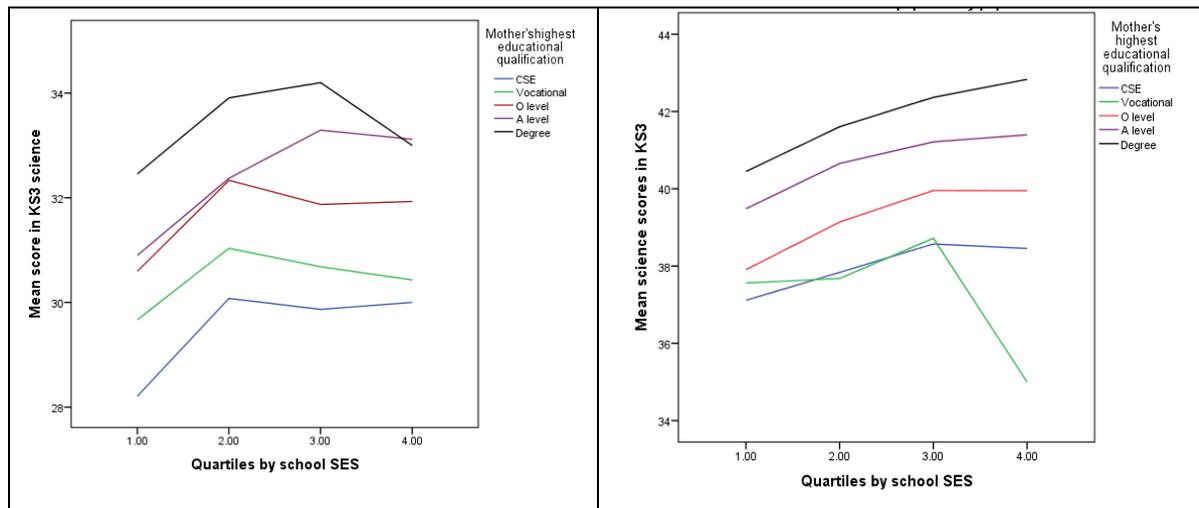
In summary, the analyses of KS2 and KS3 science attainment are consistent with the findings in the international literature: school SES is possibly a mediator, and not a moderator, of the impact of SES on science attainment. The analyses that follow concentrate on individual differences rather than differences between schools.

b. **Reading comprehension, scientific reasoning and vocabulary as mediators of the impact of SES on science attainment**

The large body of data in the ALSPAC includes information about the participants' performance on measures of relevance for analysing what mediates the impact of SES on science attainment. Our analyses focus on the factors identified as potential mediators of this connection in chapter 4, namely science reasoning and reading comprehension, and on vocabulary as a proxy for language skills, in order to test the validity of teachers' beliefs that pupils from lower SES backgrounds struggle with science due to poor language skills. Although self-regulation seemed a plausible mediator of the impact of SES on science attainment on the basis of the literature review, the measures included in ALSPAC did not meet the criteria for being a possible mediator of the SES-science attainment relation, and thus are not considered here.

In this section we describe several regression analyses designed to investigate to what extent the inclusion of reading comprehension and scientific reasoning in a regression equation reduces the impact of SES on science attainment. The analyses of science attainment are reported by key stage tests.

The main way of testing whether or not the three variables - reading comprehension, scientific reasoning and vocabulary - are mediators of the relation between SES and science attainment was to compare the results of two or more regression analyses. For example, in one analysis we looked at how much of the variance in pupils' science attainment is accounted for, or explained, by a single predictor, SES. In the next analysis we entered two predictors - for example, SES and reading comprehension - to see how well they predict the same outcome measure. If the introduction of reading comprehension into the regression reduces the amount of variance in pupils' science scores explained by the their individual SES, it would be reasonable to conclude that part of the reason for the relation between SES and science attainment is due to the effect of differences in reading comprehension. In other words, the result would mean that reading comprehension partially mediates of the relation between SES and pupils' science learning.

e. Science attainment, reading comprehension and vocabulary

The first step in the analysis of possible mediators is to look at the amount of variance explained by SES in science attainment at KS2 and KS3. On its own, SES accounted for 14.8 % of the variance in the KS2 science results[16], for 6.9% of the variance in the KS3 results for paper A and 10.5% of the variance in paper B.

Let us now look at reading comprehension, which is the literacy measure that has most often been shown to predict pupils' science learning. To check whether reading comprehension is a mediator in the SES-science learning relation, we ran a multiple regression with science attainment as the outcome measure and two predictor variables, reading comprehension (entered first) and SES (entered second). If the amount of variance explained by SES in science attainment decreases in this analysis, this must be due to our entering reading comprehension into the equation first, and the consequent decline in the SES-science relation must be due to the mediating role of reading comprehension.

In fact, the two-step regressions in which we entered reading comprehension first and SES second produced this pattern of results with KS2 and KS3 science attainment. The scores in the reading comprehension test, which was administered when the pupils were about 9-years old, accounted for a whopping 37.6% of the variance in their KS2 science test scores while SES accounted for 2.2% of the variance in these outcome scores after the impact of reading comprehension was taken into account – much smaller than when this variable was the only predictor in the equation, but still significant (p<.0001) because of the large number of participants.

Turning to the KS3 science results, we find a very similar pattern in both science papers. In paper A, reading comprehension accounted for 31.2% of the variance in the science scores 5 years later, while SES accounted for 1.30% of this variance. Although this is still a significant result, it is very much smaller than the 6.9% of the variance in science scores explained by SES in KS3/Paper A attainment when this was the only predictor variable. In paper B, the two-step multiple regression showed that reading comprehension accounted for 25% of the

---

[16] There are small variations in the amount of variance explained across analyses because different numbers of participants are included. In the previous analyses of KS2 outcomes, only those children who had attended the same school in years 4 and 6 were included. This restriction does not apply here. The restriction that applies here is that the participants must have taken the different measures included in the analyses.

variance in science attainment while SES accounted for 4%; this is a considerable reduction in the amount of variance explained by SES on its own, which was 10.5%.

In order to understand better the mediating role of reading comprehension, we then ran three-step multiple regressions in which reading comprehension and vocabulary were entered as predictors. It is known that reading comprehension and vocabulary are highly correlated; it could be the case that the effect of reading comprehension is actually due to its relationship to vocabulary. Because teachers conjecture that children from lower SES backgrounds are at a disadvantage in learning science due to their lower language skills, we thought that it was reasonable to test whether the impact of reading comprehension on science attainment was itself mediated by the children's language skills. This required running first a three-step analysis, in which vocabulary was entered first in the equation, then reading comprehension, and finally SES as predictors of KS2 and KS3 science attainment. It was sensible to run another three-step analysis in which reading comprehension was entered first and vocabulary second, as this would allow us to find out the amount of variance in science attainment explained independently by reading comprehension and vocabulary. The amount of variance explained by SES after these two measures are entered in the equation does not vary with the order in which reading comprehension and vocabulary are entered.

In the analysis in which we entered vocabulary first and reading comprehension second, vocabulary explained 26.5% of the variance in KS2 science attainment and reading comprehension explained a further 15.7%; SES now only explained 1.1% of the variance in KS2 science attainment. This means that the effect of reading comprehension was not entirely attributable to its connection to vocabulary, as it still explained a large amount of variance when vocabulary had already been entered into the equation. When reading comprehension was entered first in the equation, it explained 37.6% of the variance, as noted earlier on, while vocabulary only explained a further 4.6% of the variance in KS2 science attainment, which is a very large reduction from its explanatory power on its own, but still significant. Thus the analyses show that the effect of reading comprehension on science attainment is not explained by its correlation with vocabulary, as it accounts for a large amount of variance when vocabulary has been taken into account. Reading comprehension and vocabulary make independent contributions to the prediction of KS2 science attainment, but the contribution of reading comprehension is undoubtedly larger.

The pattern was very similar indeed when the three-step analyses were run for KS3 science attainment. In Paper A, when vocabulary was entered first, it explained 20.6% of the variance and reading comprehension explained a further 16.0%. Although this is a reduction in the amount of variance it explained on its own, which was 31.2%, it still explains a considerable amount of variance in Paper A, which the participants took 5 years after the reading comprehension test. When reading comprehension was entered first, vocabulary only explained 6.8% of the variance, which is still a considerable amount but only about one third of the amount of variance it explained on its own. When both reading comprehension and vocabulary were entered in the analysis before SES, the latter only accounted for 0.6% of the variance in KS3/Paper A.

The pattern is much the same for Paper B: vocabulary accounted for 15.2% of the variance in KS3/Paper B and reading comprehension accounted for 12.9% when entered second. This is a reduction in the amount of variance explained by reading comprehension when it was entered first, which was 24.8%, but reading comprehension still plays a large part in

explaining science attainment in Paper B, after taking into account the effect of vocabulary. Once again, the amount of variance explained by vocabulary is reduced quite considerably after taking into account reading comprehension, as it only explains 3.2% of the variance in science attainment. The effect of SES was reduced to explaining 2.6% of the variance in these analyses. Figure 5.4 summarises these results.

Figure 5. 4 A summary of the one step regression (top) and two steps regression (bottom) carried out to test whether reading comprehension mediates the SES-science attainment relation. The percentages indicate amount of variance explained in the KS assessments. The number by the arrow connecting SES and reading comprehension shows the correlation between these two measures.
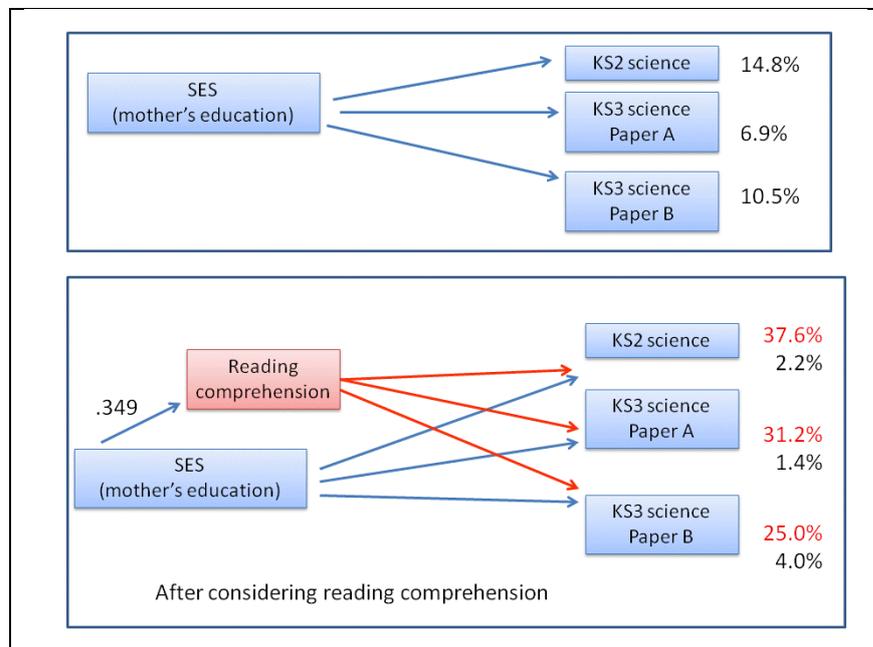
In summary, analyses of science attainment in KS2 and in both KS3 papers support the hypothesis that reading comprehension is a mediator of the impact of SES on science attainment. In spite of the moderate correlation between reading comprehension and vocabulary, both of these measures explain independent amounts of variance in KS2 and KS3 science outcomes; reading comprehension consistently explains more variance than vocabulary. The assessments of reading comprehension were given to the participants when they were between 8 and 9 years; KS2 science is given at 11 and KS3 science at 14 years. The findings are therefore replicated with different measures of science attainment, given 2 and 5 years after the participants had taken the reading comprehension test. The number of participants in the study is always rather large and the measures are all standardised. These analyses provide very strong support for a mediating role for reading comprehension in the SES-science attainment relation and also a robust, but not as strong, role for vocabulary. It is noted that, when these two predictors are entered together in the analyses, the SES-science attainment relation becomes very weak indeed.

It is worth pausing to reflect on why reading comprehension is a stronger predictor of science attainment than vocabulary. First, and most obviously, reading comprehension involves vocabulary, but also syntactic knowledge, because sentences are composed with words but their organisation in sentences adds much meaning to the information conveyed by words. Second, reading comprehension requires the reader to process larger chunks of information, and thus makes higher demands on processing skills. Finally, although most often it has been assumed that children's vocabulary knowledge is one of the variables that explains their reading comprehension, there is research that suggests that this is a reciprocal relation, and that reading comprehension actually explains much about vocabulary knowledge. Smith (1941) and Nagy and Anderson (1984) examined vocabulary growth in primary school pupils. Their estimate is that pupils' vocabulary growth is on average of more than 5,000 words a year as they go through primary school. Nagy, Herman and Anderson (1985) suggest that this amazing acquisition happens without much help from teachers. Teaching children this number of words one by one is a hopeless enterprise. They therefore tested the hypothesis that children learn many words incidentally from reading, because the number of words used in oral language is a small sample from the number of words printed in primary school books. They did, in fact, find that exposing children to texts that contained new words promoted word learning, an accomplishment that depended on their reading comprehension (i.e. their ability to infer word meanings from text) and their morphological knowledge (Nagy & Anderson, 1984). As discussed in Chapter 4, many words used in science do not occur frequently in everyday conversations; children from all SES backgrounds might need explanations of their meanings, or might have to learn them from context and their knowledge of morphology. It is thus understandable that, in spite of the significance of learning new vocabulary for science learning, reading comprehension can be an asset in this process, rather than simply an outcome of vocabulary knowledge.

f. Science attainment and scientific reasoning

Let us take the same steps for testing whether another predictor variable, scientific reasoning, can be a mediator of the link between SES and science. In three further multiple regressions we entered the pupils' scientific reasoning scores measured by the Control of Variables task (administered when the pupils were 11-years old) as the first step and SES as the second step in the regression. Again the outcome measures were the KS2 science scores in one analysis and the KS3 Papers A and B science scores in the other two. In the first of these analyses the scientific reasoning scores accounted for 24% of the variance in KS2 science scores, which is an impressive figure, and SES accounted for 5.9% of the variance in these outcome scores, which is also impressive, but smaller than when SES is entered as a predictor on its own.

In the second analysis, with KS3 Paper A as the outcome measure, scientific reasoning accounted for 9.2% of the variance in science attainment while SES accounted for 4.5%, a reduction of 2.4% in the variance accounted for on its own. The total amount of variance explained increases from 6.9%, when only SES was used as a predictor, to 13.7% when scientific reasoning is included in the equation, which indicates that the variance explained by scientific reasoning overlaps only partially with the variance explained by SES.

In the third analysis, with KS3 Paper B as the outcome measure, scientific reasoning accounted for 14.9% of the variance in the participants' KS3 science scores, while the contribution of SES was found to account for 7% of the variance, which is smaller than when

SES was the only predictor variable and accounted for 10.5% of the variance in KS3 Paper B, but not as large a reduction as when reading comprehension is treated as the mediator. The total amount of variance explained by scientific reasoning and SES is 21.8%, about twice the amount explained by SES on its own, which reinforces the conclusion that the variance explained by scientific reasoning and SES overlaps, but only partially. Remember that the scientific reasoning test is a longitudinal predictor of KS3 science attainment, as this assessment was given to the participants about three years before they took the KS3 science test. This is therefore an important result. Figure 5.5 presents the results of these analyses at a glance.

Figure 5. 5 A summary of the one step regression (top) and two steps regression (bottom) carried out to test whether scientific reasoning mediates the SES-science attainment relation. The percentages indicate amount of variance explained in the KS assessments. The number by the arrow connecting SES and scientific reasoning shows the correlation between these two measures.
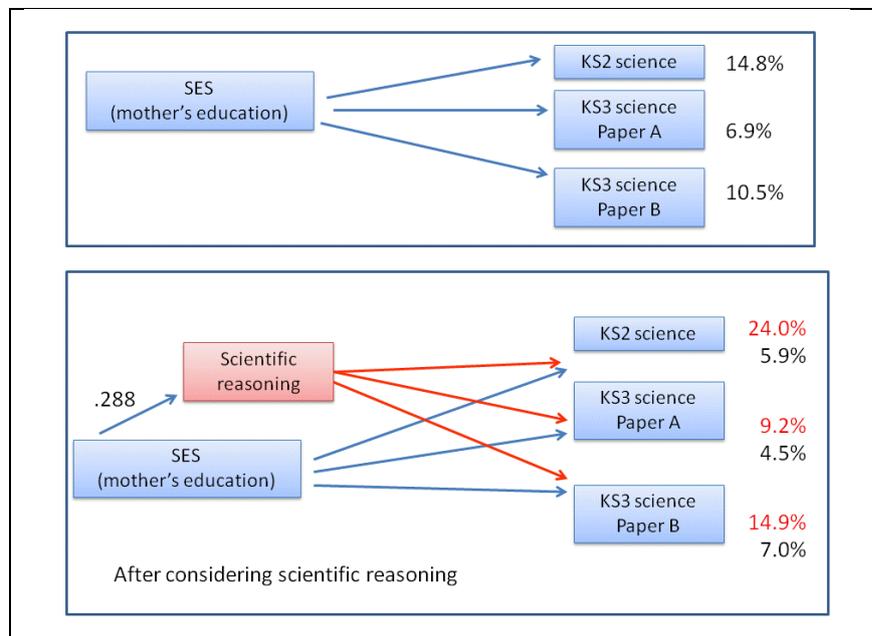
To summarise, the three candidates won their mediator status. Pupils' reading comprehension, their vocabulary and their scientific reasoning are strongly related to their progress in science and to SES. Each variable accounted for a large amount of the variance in pupils' science attainment longitudinally; reading comprehension and vocabulary were separated from KS2 by about 2/3 years and from KS3 by about 5/6 years and scientific reasoning was separated from KS3 by about 3 years. The pupils' SES levels also predicted these two sets of KS science attainment scores but, when the reading comprehension skills, their vocabulary knowledge, or their scientific reasoning skills were entered as the first step in the equation and SES as the second step, the relationship between SES and the pupils' science attainment was much reduced. This reduction is a sign that we have established the existence of three important mediators of the connection between SES and science learning.

g. Science attainment, reading comprehension and scientific reasoning

Having established the existence of three strong mediators, and that reading comprehension and vocabulary made independent contributions to predicting science attainment but that reading comprehension was the stronger predictor of the two, our next step was to see whether reading comprehension and scientific reasoning are in any way independent of each other in their mediating role. We carried out three three-step multiple regressions; in one the outcome measure was KS2 science scores and in the other two KS3 science scores, one for each paper. The first variable that we entered in each analysis was the pupils' reading comprehension, the second their scientific reasoning and the third their SES. In the first analysis the outcome measure was KS2 science scores and in the other two KS3 science scores, one analysis for each paper. These analyses were designed to answer two questions: (1) whether scientific reasoning scores would predict any additional variance in the science outcomes over and above the variance already explained by the pupils' reading comprehension scores, and (2) how much of the variance in the science scores was still due to SES differences after we had controlled for the impact of both these two powerful mediators.

The answer to the first question was that the pupils' science reasoning scores continued to predict the KS 2 and 3 science scores even when it was entered as the second step, after reading comprehension. When the KS2 science results were the outcome measure, the scientific reasoning scores explained as much as 7.4 % of the variance after the impact of reading comprehension had been taken into account. When the KS3 Paper A science results were the outcome measure, scientific reasoning still explained 2.6% of the variance in science attainment. In KS3 Paper B, the variance explained by the pupils' scientific reasoning scores (measured when the pupils were 11-years old) accounted for 6.7% of their science results at 14 years. So, even over a gap of three years, the pupils' scientific reasoning continued to make an independent prediction of variation in science attainment scores. We conclude that the two mediators make relatively independent contributions to pupils' learning of science in their first years at secondary school.

The answer to the second question is quite remarkable. Remember that on its own SES accounts for 14.9% of the variance in the KS2 science results. We have seen that this contribution by SES to the prediction of the KS science attainment was much reduced when each of the mediators was entered into the equation separately in two different two-step analyses. Now, the three-step analysis, which includes both mediators, reduces the impact of SES on the science scores even further. In this multiple regression SES accounted for only 1% of the variance in the pupils' KS2 science attainment. The lion's share of SES differences in science learning depends on the working of the two powerful mediators – reading comprehension and scientific reasoning.

Remember also that in KS3, SES on its own accounted for 6.9% of variation in pupils' KS3 attainment in Paper A and for 10.9% in the Paper B. After reading comprehension and scientific reasoning scores were taken into account, SES accounted for only 0.8% of variation in pupils' KS3 attainment in the Paper A and 2.1% in the Paper B. Figure 5.6 presents these results at a glance.

Figure 5. 6 A summary of the one step regression (top) and the three steps regression (bottom) carried out to test whether reading comprehension and scientific reasoning mediate together the SES-science attainment relation. The percentages indicate amount of variance explained in the KS assessments.
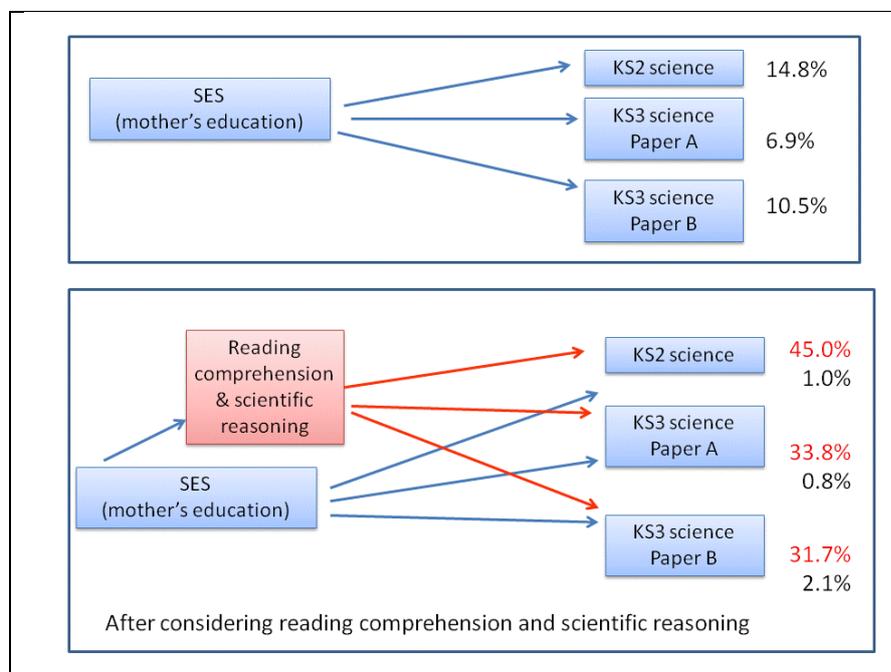
The relation between SES and science attainment: is it all about IQ?

One of the points that we made earlier on with respect to the connection between SES and science attainment was that there is a possibility that there is no causal relation between the two, but that they are both explained by a third factor, namely intelligence. Intelligence tests were designed to predict educational outcomes and are validated by their ability to do so. Thus their ability to predict educational outcomes on their own is not very informative; in fact, it is a circular argument to conclude, after validating intelligence tests by correlating their results with educational outcomes, that intelligence explains educational outcomes.

Cattell (1963), in his classic theory of intelligence, argued for a theory of general abilities that distinguishes two factors, "fluid and crystallized general abilities" (Cattell, 1963, p.1). Crystallised ability is measured by tasks in which previously learned problem solving "habits" (his expression) are advantageous; fluid intelligence is measured by novel tasks, in which the

advantage from previously learned habits is less important. Although Cattell recognises that the two factors operate together, he argued that it was possible to separate them to some extent through factor analysis. He exemplified crystallised intelligence as measured by tests of verbal or numerical ability as well as attainment in history and geography; science attainment might have been classified with these other attainments but he did not cite it. Cattell also indicates that the sort of intelligence studied by Piaget in his reasoning tasks is part of crystallised intelligence; he sees the sort of problem solving structures that Piaget investigated as "aids" for solving problems and learning.

Since Cattell's work, many other researchers have incorporated this distinction between crystallized and fluid abilities into their models of intelligence (e.g. Baddeley's 2000 model of working memory includes a component, the episodic buffer, which relates working memory to long term memory, i.e. what has been learned). The measure of intelligence used in the ASLPAC, the WISC, is traditionally viewed as made of two components, verbal and non-verbal intelligence, but a principal component analysis identifies three, as described previously: verbal comprehension, perceptual organisation, and freedom from distractibility. In the analyses we are about to describe, we decided to include two factors as predictors: perceptual organisation, and freedom from distractibility, and not to include verbal comprehension because verbal comprehension is related to vocabulary and reading comprehension (see, for example, Hoover & Gough, 1990, whose theory was adopted by the Rose Review, 2009, as the foundation for the literacy curriculum in the UK). Thus the previous analyses have already considered the role of verbal ability in explaining the link between SES and science attainment as measured by reading comprehension and vocabulary. The two factors, perceptual organisation and freedom from distractibility, met the criteria described earlier on in this chapter: each correlates significantly with SES, KS2 and KS3 measures of science attainment.

Initially we ran three simple regressions to assess how much variance in KS2 and KS3 science attainment was accounted for these two factors. We entered them into the analyses in a single step, as we have no reason to treat them differently. These two factors on their own explained 26.3% of the variance in pupils' scores in KS2 science, 31.0% of the variance in KS3 Paper A, and 20.5% of the variance in Paper B. These are substantial amounts of variance accounted for in each of the three measures of science attainment.

In order to ascertain whether the pupils' scientific reasoning and reading comprehension make an independent contribution to the prediction of science attainment, above and beyond the contribution from perceptual organisation and freedom from distractibility, we ran three regressions in which these two factors were entered first, and scientific reasoning and reading comprehension were entered afterwards. In each of the three analyses, scientific reasoning and reading comprehension together explained further variance. In the KS2 analysis, reading comprehension and scientific reasoning accounted for a further 13.5%, after controlling for perceptual organisation and freedom from distractibility; the amount of variance explained in the analysis increased to 39.7%, which demonstrates that a large proportion of the shared variance between reading comprehension, scientific reasoning, and science attainment is not accounted for by perceptual organisation and freedom from distractibility. In KS3, the further amount of variance explained by reading comprehension and scientific reasoning, after controlling for the two factors from the WISC, was 13.5% for Paper A and 16.4% for Paper B. The amount of variance explained increased to 41.5% and 37.3% when reading comprehension and scientific reasoning were entered in the analyses.

It thus seems safe to conclude that the impact of reading comprehension and scientific reasoning on science attainment is not explained by perceptual organisation and freedom from distractibility.

In contrast, when reading comprehension and scientific reasoning were entered in the hierarchical regression model first as a single step, and perceptual organisation and freedom from distractibility were entered second, also as a single step, the first two factors explained 36.2% of the variation in pupils' science attainment in KS2 while the two factors from the WISC measure explained 3.8%. This is a considerable reduction in the explanatory value of the measures of perceptual organisation and freedom of distractibility, which suggests that they cannot explain the connection that reading comprehension and scientific reasoning have with science attainment.

Similar results were obtained in the analyses of KS3 science attainment. In KS3 Paper A, reading comprehension and scientific reasoning explained together 32.9% of the variance and the two factors from the WISC explained 9%, a considerable reduction from the 31% of the variance explained when the factors had been entered in the regression equation on their own. In KS3 Paper B, reading comprehension and scientific reasoning explained together 32.2% of the variance and the two factors from the WISC explained 5.1% when entered as a second step, again a considerable reduction from the 20.5% that they had explained on their own.

In conclusion, the connections of scientific reasoning and reading comprehension with science attainment cannot be explained by measured intelligence. The implication of this very clear result is that there is much room for schools to improve pupils' science attainment, as both reading comprehension and scientific reasoning have been found to be susceptible to teaching interventions.

Table 5.2 summarises the nature of the connection these two variables with SES and science attainment in the ALSPAC data.

Table 5.2 Summary of the relations of scientific reasoning and reading comprehension with SES and pupils' science attainment in the ALSPAC study

Predictive tasks	Relation between scores in the predictive task & SES	Longitudinal prediction of science attainment by the predictive task	Reduction in variance explained by SES when the mediator is included in the analysis and control for extraneous variables
Reading comprehension and vocabulary	The relation between SES and reading comprehension was significant in the ALSPAC study. (see Table 5:1).	The ALSPAC results confirm the common finding of a strong relation between pupils' reading comprehension scores and their attainment in science. This study, being longitudinal, also establishes that this is a predictive relation (see Table 5:1).	Including reading comprehension in the prediction of KS science attainment greatly reduced the magnitude of the SES effect. The effect was further reduced when reading comprehension and vocabulary were entered into the equation before SES.
Scientific reasoning – as measured by the Control of Variables task (CVT)	Pupils' SES was significantly related to their CVT scores in the ALSPAC study: (see Table 5:1).	The pupils' CVT scores predicted their key stage (KS) scores in science, both in KS2 science taken in the same year as the CVT task and in KS3 science taken 3 years later on (see Table 5:1).	Including scientific reasoning in the prediction of KS science attainment greatly reduced the magnitude of the SES effect.

Reading comprehension and scientific reasoning	The joint effects of scientific reasoning and reading comprehension on KS science attainment did not disappear when measured intelligence was entered into the equation before these measures. The effect of measured intelligence was greatly reduced when this measure was entered after reading comprehension and scientific reasoning.

Conclusions and discussion

- The analyses of the ALSPAC data showed that the SES-science attainment link is to some extent mediated by institutional factors, such as school SES, and to a larger extent mediated by individual factors, such as reading comprehension and scientific reasoning. When both reading comprehension and scientific reasoning are taken into account, the effect of SES on science attainment is negligible. The mediating role of reading comprehension and scientific reasoning cannot be explained by the students' measured abilities.

- What are the educational implications of these results? The educational point here is that, since so much of the SES differences in science attainment depends on pupils' reading comprehension and on their scientific reasoning, improvements in both these skills should narrow the gap between students from different SES levels in learning about science. Improve students' reading comprehension and their scientific reasoning and the gap should diminish or perhaps disappear altogether.

- This is a testable hypothesis. Evidence reviewed in the previous chapter shows that it is possible to improve children's reading comprehension in the context of science (Morrow, Pressley, Smith & Smith, 1997) and also their scientific reasoning (Klahr & Nigam, 2004: Zohar and Peled, 2008) through specific teaching targeted at these forms of knowledge. We predict that an intervention programme that combines the teaching of both skills will help children to learn science and will diminish the regrettable, widespread and highly consistent differences that now exist in the science attainment of students from different socio-economic backgrounds.

Chapter 6

A review of current evidence on promising educational approaches that are likely to improve the attainment and progression of low-SES pupils in science education[17]

Box 6 Chapter 6 Summary

A systematic and rigorous review of the research literature has been conducted to gather and evaluate the evidence for promising educational approaches that are likely to improve the attainment and progression of low-SES pupils in science education. The main findings are as follows:

- Teaching programmes that aim to improve pupils' scientific reasoning or their literacy (variables cited as possible mediators in earlier sections of the report) have been generally successful. The success of these programmes depends partly on the provision of effective professional development, needed to ensure high fidelity in the interventions. So far, research on these programmes has been largely confined to pupils in upper primary or early secondary school years. The programmes focused on literacy aimed to develop pupils' literacy in scientific reasoning, suggesting that there may be scope for multi-faceted programmes.

- The number of studies on teaching programmes designed to improve low-SES pupils' metacognitive skills is low, but the existing research suggests that interventions designed to develop pupils' group work skills, and to teach them to evaluate and make use of their own assessment data are beneficial to their learning of science. Again, this research focuses on pupils in upper primary or lower secondary school.

- Studies of the socio-cultural aspects of teaching science have concentrated on three issues:
  1. Bringing pupils into a science 'place' e.g. university laboratories or a science museum
  2. Bringing scientists or extra-curricular science activities into schools
  3. Developing teachers' understanding of pupils' perspectives

This research supports the hypothesis that residential fieldwork and the use of informal science education institutions improves pupils' learning of science, provided that these experiences are set up in carefully structured ways. The evidence from other studies suggests that there are benefits of after-school activities, such as STEM clubs, and peer-mentoring. Socio-cultural studies also suggest that teachers can be helped to develop more culturally relevant pedagogy. Many of the socio-cultural studies are with pupils older than 14 years.

---

[17] This chapter was prepared by Judith Hillier and Jaimie Miller-Friedmann

Introduction

This final chapter of the review addresses the question of what evidence exists to support the claims made about interventions aiming to improve the attainment and progression of low-SES pupils in science education, and the use of this evidence to propose future promising avenues of research in this field. The systematic review was tightly focused on interventions, (not exploratory work), with school-aged low-SES pupils in science since 1990, with preference given to robust research designs. The details of the methods used and the criteria applied in the search are presented in Appendix 6.1. Although much science education research has been published internationally and in the UK (e.g. summaries by Wellington & Ireson, 2008; Osborne & Dillon, 2010), it was disappointing to find that relatively little work has been done within the parameters we have described, and almost none of it conducted within the UK. Indeed research carried out in the USA dominates the discussions. Nevertheless, we have found evidence for a range of different educational approaches, including the development of pupils' scientific reasoning and inquiry skills, their literacy and metacognitive skills, and research drawing on the socio-cultural perspective. The studies reviewed identified a number of features of successful interventions and, together, these indicate potential for promising further work.

There are clear links between pupils' success in learning science and both their scientific reasoning ability and their literacy levels, which were identified as possible mediators in Chapters 4 and 5 of this report, and with metacognitive ability which was categorised as a possible mediator in Chapter 4. Several intervention studies in each of these areas produced positive results (though not always with a reported effect size). There are also some studies within the socio-cultural framework which aim to break down perceived barriers between pupils and studying science, again reporting positive results, though not always effect sizes. Given this agreement with previous chapters, it seemed obvious to use these categories to structure the presentation of the findings. Table 6.1 lists the numbers of the studies that fell into the resulting categories, and also shows the age groups of the pupils involved in the interventions (these groups correspond to the Key Stages in the English National Curriculum). Note that some studies fell into more than one category; for example Oliver et al. (2012) aimed to develop pupils' scientific reasoning and metacognitive skills, and Kaldon & Zoblotsky (2014) worked with pupils aged 6-13 years, so falls across three age groups. This made categorisation of the aim of the interventions awkward, but emphasises the complex and multi-faceted nature of the problems facing the teaching and learning of science for low-SES pupils. The studies of pupils' scientific reasoning skills will be discussed first, as this contains the largest number that reported effect sizes. The emphasis will be on the studies which did report effect sizes, though other promising studies will be brought into the discussion as additional examples. Appendix 6.2 provides a summary of those studies which present figures for effect sizes. For each study, the socio-economic status of the pupils will be reported in the same way as in the original paper.

Table 6. 1 Focus of science education research papers reviewed

Focus	Number of studies	Number with effect sizes	Range of effect sizes	Number based in UK	Number of interventions longer than 1 month	Age of pupils				
						5-7 years	7-11years	11-14 years	14-16 years	16-18 years
Scientific reasoning and inquiry	35	11	0.02-1.84	1	16	11	13	13	5	2
Literacy	12	3	0.3-0.52	0	7	3	4	3	2	0
Technology	11	2	0-1.94	0	6	0	4	4	3	3
Metacognition	4	3	0.24-1.09	0	4	0	2	2	2	0
Socio-cultural	49	0	-	5	30	3	7	17	14	10

a. Studies aiming to develop pupils' scientific reasoning and inquiry skills

Developing pupils' scientific reasoning and inquiry skills was the most common aim of the studies which reported effect sizes (11 out of 15). However, there was considerable variation in the design of the studies and in the strength of the evidence presented. Typically, a curriculum programme was introduced, accompanied by professional development for the teachers implementing the programme. Indeed, a lack of professional development provision for teachers produced a negative effect on pupil attainment in one study (Doppelt et al., 2009). Pre- and post-testing was utilised, in a random control trial (RCT) design by three of the studies (Griggs et al., 2013; Hand et al., 2016; Tong et al., 2014), and another used randomised pair-matching (Kaldon & Zoblotsky, 2014). The rest involved a range of case study designs. The studies were conducted with pupils of various ages:

- Six of these were with primary school age pupils (Diaconu et al., 2012; Griggs et al., 2013; Hand et al., 2016; Kaldon & Zoblotsky, 2014; Ruby, 2006; and Tenenbaum et al., 2004)
- Five were with Key Stage Three age pupils (11-14 years) (Doppelt et al., 2009; Marx, et al., 2004; Oliver et al. 2012; Thomas et al, 2015; and Tong et al., 2014)

This appears to reflect the often expressed view in the literature that work to ameliorate the negative effects of low-SES on pupils' attainment should begin earlier in their educational journey rather than later (e.g. Tong et al., 2014). Although much research has been carried out in this field, there are a number of definitions for 'inquiry' in science education. One review paper offered the following list of features:

"questioning and generating hypotheses, experimenting, designing, and planning, predicting, modelling/ visualizing, observing and data collection, analyzing data, interpreting and explaining, developing/evaluating/arguing, reaching conclusions, and communicating findings." (Donnelly et al., 2014, p.573)

The strength of the evidence presented in these articles varies considerably. Not all studies measured pupil attainment directly: some used quite narrow assessment instruments, some had relatively small sample sizes, and a range of effect sizes is reported. As such, the studies

can be ordered along a continuum, from least convincing to most convincing, and will be presented in this order.

Diaconu et al. (2012) measured a significant improvement in teachers' scientific content knowledge, teachers' use of classroom strategies (self-reported) and teachers' leadership skills (self-reported), with an effect size of 0.1 after a one year intervention providing professional development for primary school teachers in urban schools in Houston, Texas, USA, where more than 75% of pupils, aged 8-10, were economically disadvantaged (although details of this categorisation were not given). These researchers stated that their longitudinal multi-site case study was based on the two assumptions that better teacher knowledge and more use of inquiry-based teaching practice would lead to increased pupil attainment. Although this has been found to be true in other studies reported here (e.g. Doppelt et al., 2009; Marx et al., 2004), the evidence from this study would have been strengthened by a direct measurement of pupil attainment. Nevertheless, the participants highly valued the long-term professional development opportunity provided by this intervention, and could see the impact it was having on their teaching and their pupils: '*The pupils were engaged. They were asking questions. They were arriving at their own conclusions*' (p.870), and termed it to be a '*miracle happening in poor schools*' (p.873). The positive impact of the long-term ongoing professional development will be discussed further at a later stage.

Thomas et al. (2015) reported on a multi-site case study project in urban high schools in New York City, USA, with 64%-92% of 13-15 year old pupils receiving free or reduced price school meals, where teachers and peer leaders received on-going professional development over a year in order to support peer-led inquiry-based learning in mathematics and science. Despite the careful research design and detailed analysis of data from over 500 pupils from the end-of-year state examinations, negligible effects were observed for science. The researchers argued that this was due to the weak content knowledge of the pupil peer leaders, thus providing the first of our cautionary tales about the type of interventions which are less promising.

Effect sizes were also reported in studies by Ruby (2006) (between 0.02 (not significant) and 0.2), Kaldon & Zoblotsky (2014) (0.04), and Tenenbaum et al. (2004) (0.11 and 0.24). Ruby (2006) looked at the findings of a 10 year initiative to support middle school teachers in Philadelphia schools in the USA where 71%-95% of pupils aged 8-12 years came from low income families, (although this term was not defined), with 630 pupils in the treatment group and 463 in the control. The longitudinal, non-equivalent group design had a four-strand approach including '*(1) an implementable curriculum, (2) ongoing intensive teacher professional development, (3) in-classroom support from peer coaches, and (4) mechanisms to foster and sustain changes in science instruction, including opportunities for science teacher cooperation and development of teacher leaders*' (p.1009). The significant effect sizes of 0.18 and 0.2 were seen in the schools with a high level of fidelity of implementation of the intervention, and Ruby suggested that the peer coaching was crucial to this success. He also recognised that it was not possible to disentangle the effects of the curriculum materials from the effect of the other professional support mechanisms. However, the study could not be continued further as state-wide education reform reorganised the school structure and curriculum, providing another example of the difficulties in this sort of research.

Kaldon and Zoblotsky (2014) studied another long-term (5 years) project with 60 000 pupils aged 6-13, (70% economically disadvantaged, though this term was not defined), which aimed to embed an inquiry-based science curriculum in schools through professional development for teachers, school leaders and administrators in a randomised matched-pair design. The effect size of 0.04 on pupil attainment was deemed insignificant by the researchers, and the effects on pupils' attitudes also seemed to be negligible. This paper was an interim report on the project, and we have been unable to locate a peer-reviewed final analysis, hence, it is difficult to account for the low effect size. One possibility is that, with such large numbers, problems may have arisen with fidelity of implementation, echoing the concern raised by Ruby (2006).

The project by Tenenbaum et al. (2004) was a small-scale multi-site case study, involving 30 kindergarten children (age five) in the USA (plus 18 as a control), with 72% on free or reduced price school lunches. This was one of the shortest studies reviewed, taking place over about a month during which the children visited a local children's museum and participated in two classroom lessons on the content of the science exhibits they had visited at the museum. The intervention was designed to support the children in developing their own explanations about buoyancy (effect size 0.11), bubbles and currents in water (effect size 0.24), (the researchers considered these to be medium effect sizes), and the findings showed that the children developed increased concept complexity, though they were not more accurate at predicting whether an object would float or sink. Although the authors could claim that learning had occurred, it was unclear whether it had occurred in the museum, or in the classroom, or both. This suggests two important points, one about how learning rarely occurs at a single moment in time, and the other about the need to avoid overly simplistic measures of attainment – developing better scientific reasoning as a step towards working out the right answer could be considered more progress than simply getting the answer right in the first place.

Finn et al. (2015) conducted a small-scale case study with 47 pupils aged 11 to investigate the effect of using physical exercise during after-school childcare in an urban community in Massachusetts, USA, over six weeks to support an inquiry-based approach to the teaching and learning of science, namely the physiology aspects of the biology curriculum. All the pupils were low-SES (economically disadvantaged without further explanation), with a pre- and post-test assessment of science learning and effect sizes of 1.09 for the treatment group (n=16) and 0.5 for the control (n=31). Although after-school childcare could be a valuable opportunity to introduce extra provision for low-SES pupils, the sample size here is very small, and the study relates to a narrow part of the curriculum. Further work is needed to explore what potential this might offer.

The question of how to support the development of pupils' scientific reasoning and inquiry skills is continued in the next study by Oliver et al. (2012) which examined the introduction of CASE (Cognitive Acceleration through Science Education) in rural Australian schools where 57% of pupils had parents whose income was in the lowest quartile. CASE has previously been introduced in the UK with promising effects (for more details see Adey and Shayer, 1990; Shayer and Adey, 1992a and 1992b). The Australian study was conducted with 68 pupils, aged 12-14 years, and six teachers in one school in a case study approach, with the teachers receiving six days of professional development over two years, and working collaboratively to support the teaching of the 30 CASE lessons (over the same two year period). The science reasoning tasks used as pre- and post- tests showed an increase in

cognitive gain with an effect size of 0.47, though the effect was smaller for pupil attainment in the state-wide and national tests in science (0.21). The authors have no information on whether this result would improve academic attainment at a later age and planned to research this further. The vast majority of the pupils had positive perceptions of the CASE lessons, and most teachers were positive about the intervention, citing the enjoyment of pupils and the opportunities to develop scientific reasoning skills. However, some teachers were less enthusiastic about the intervention, particularly those who had been teaching a long time, and the success of the intervention had clearly depended on the support and involvement of the school's head teacher, again points to note for the later discussion. Finally, the authors were careful not to overstate the effect CASE could have on low-SES pupils: clearly an intervention with potential, but not straightforward in implementation. These findings are reflected in a recent UK-based adaptation of CASE which showed no evidence of improvement in pupils' attainment in science, but which appeared to have poor fidelity of implementation (Hanley et al., 2016).

A large multi-site case study was conducted in the USA by Doppelt et al. (2009) on the effects of introducing a design-based curriculum unit, which took the form of an open-ended project focused on engineering and the use of scientific reasoning in the context of electricity, supported by professional development over two years. Here five teachers and 405 pupils formed a control group, five teachers and 274 pupils participated in the curriculum reform, as summarised above, but not the professional development, and 13 teachers and 977 pupils participated in both. Between 70% and 84% of the pupils, aged 12-14 years, qualified for free or reduced price school lunches. The effect of introducing the curriculum reform without the professional development was negative (effect size 0.5), but the effect of both the curriculum reform and the professional development was positive (effect size 0.67). During the first year, pupils' understanding was assessed through six core questions about electricity. This was expanded to 20 questions during the second year, with little change to the findings. The professional development consisted of five four-hour workshops spread over the course of the curriculum unit, which is a significant amount of time for a teacher. However, the authors felt this length of time was crucial to the success of the intervention as it enabled the formation and development of a professional learning community where teachers were highly motivated to attend the workshops because of the impact they felt these had on their professional practice. The control group was taught the same electricity content through what the authors described as 'scripted inquiry learning', which is perhaps another cautionary note about how the previous well-intentioned curriculum reform had stagnated in its implementation.

The final study that focussed on low-SES pupils, which is to be discussed in detail in this section, is by Marx et al. (2004) who analysed the results of a three year project in Detroit urban schools in the USA, (~50% pupils from families at or below the poverty line). Again, this comprised multiple strands to science education reform: introducing scientific inquiry curricula where pupils made use of a range of software, supported by professional development and with emphasis on collaboration. About 8000 pupils participated across 14 schools. Effects were positive and increased over time, with the final year producing effect sizes of 1.42 (pupils aged 10-11), 1.84 and 0.70 (pupils aged 11-12) and 0.84 (pupils aged 12-13). The researchers state these effect sizes were large *because the work was embedded in a systemic reform context* (p.1073). This appears to be a carefully designed RCT-type intervention with detailed assessment of pupils' attainment and, as such, could be

considered the strongest evidence in this section for promising pedagogies and interventions to support science learning for low SES pupils. The authors' conclusion was that the prerequisites for success comprised:

'A combination of carefully designed curriculum materials, learning technologies that are embedded in the materials and serve the needs of learners, high quality professional development, and policies that support reform are necessary. But equally important, collaboration among partners is fundamental.' p.1075.

However, a note of caution must be sounded here about the claims made for learning technologies, as the age of this study means that the 'new' technologies to which it refers will have been surpassed and replaced by a great many more.

As Table 6.1 shows, approximately a quarter of the papers reviewed have scientific reasoning as a focus, usually through promoting inquiry-based learning. Our search also retrieved a small number of meta-analyses, with three of these reviewing the research published on inquiry-based learning in science education. Furtak et al. (2012) reviewed 37 studies from 22 papers and found a mean effect size of 0.5, with this increasing to 0.65 for teacher-led inquiry-based learning, and decreasing to 0.25 for pupil-led inquiry-based learning. Lazonder and Harman (2016) conducted a later review, looking at 72 studies, and found an effect size of 0.44-0.88. They also raised the question of the nature and quantity of guidance given to pupils during inquiry-led learning interventions, asking when do pupils need '*adequate guidance*' and when should they be given '*highly specific guidance*'? Donnelly et al. (2014) looked at computer-based inquiry learning research, finding an average effect size of 0.87 from their 44 papers.

Another systematic review was conducted by Slavin et al. (2014) on effective science education in elementary schools, including studies lasting at least four weeks, which used randomised or matched control groups, and measured achievement separately. However, there was no focus on low-SES pupils here. This review process resulted in 23 studies which fell into two groups: those evaluating inquiry-based teaching approaches and those evaluating use of technology to support teaching and learning. The seven inquiry-based studies which used science kits did not show a significant effect (effect size 0.02), but the 10 inquiry-based studies where there was professional development (without kits) had an effect size of 0.36. The six technology-based studies showed an effect size of 0.42. They 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.*" (p.870).

One final study will be discussed in this section on scientific reasoning and inquiry to reflect the emphasis on interventions with younger pupils. Hanley et al. (2015) conducted a review of the *Thinking, Doing, Talking Science* programme developed by Science Oxford and Oxford Brookes University, which aimed to make primary school science lessons '*more practical, creative and challenging*' (p.3). 655 Year 5 pupils from 21 schools took part in the study, with another 20 schools as the randomised comparison group, all in the UK. The intervention had a positive impact on attainment in science (effect size 0.22), with Free School Meals pupils making greater progress (effect size 0.38). The main part of the intervention was 5 days professional development for teachers and 2 days in school to plan and share resources and ideas. This adds further weight to the finding, common across

many of the studies reviewed in this chapter, about the value of professional development for teachers. Importantly, this professional development does need to be subject-specific: previous research has shown subject-specific professional development to have effect sizes of 0.4 or higher, whereas professional development about generic 'good teaching' had effects less than 0.2 (Doppelt, 2009).

Although none of these looked at low-SES pupils per se, these findings suggest that this teaching and learning approach to science education is effective for all pupils. Further research on how to promote the development of scientific reasoning and inquiry skills in low-SES pupils needs to be done.

Summary of research about scientific reasoning and inquiry skills

To summarise, the literature reviewed thus far suggests that the science attainment of low SES pupils can be increased by interventions which seek to develop their scientific reasoning and inquiry skills. Typically, successful formats include a curriculum reform or initiative which can fit into/around current curricula, accompanied by teaching materials and resources, and ongoing professional development to support teachers' use of these. Over-reliance on pupils themselves as facilitators should probably be best avoided. The professional development aspect appears crucial to enable teachers to balance the need to implement the intervention with high fidelity, but also interpret it for their own particular context. Time is needed for these interventions to become embedded into practice and for effects to be seen, and hence they are vulnerable to changing school curricula and structures rendering interventions apparently obsolete or irrelevant. They are also vulnerable to a lack of support from within schools from senior managers and administrators. This is also supported by the wider literature on science education. Finally interventions with primary and lower secondary school pupils can have promising effects, but there seems to be little research conducted with older secondary school pupils.

b. Studies aiming to develop pupils' literacy skills

Of the studies produced by the search whose aim was to develop low-SES pupils' literacy skills, three reported effect sizes: Hand et al. (2016), Tong et al. (2014) and Cromley et al. (2013). All three were RCTs that used a curriculum initiative, supported by professional development for teachers, to support pupils' literacy in scientific inquiry. In contrast to the studies reported in the preceding section, the two former studies aimed to develop pupils' scientific reasoning, whilst also developing their language skills in order to express their thinking more clearly, whilst the latter focused on pupils' comprehension of diagrams which are a specific aspect of literacy in science. It should be noted that literacy-focused science education research is conducted less widely than scientific reasoning and inquiry research (see Table 17), and we found fewer papers studying this. Also to be noted here is the difference between 'literacy in science', which is focused on language in science, and 'scientific literacy', the definition of which is discussed in Appendix 1.2.

Hand et al. (2016) worked with 32 teachers and 780 elementary school pupils (aged six-nine) over three years to embed a science writing heuristic approach which aimed to 'provide pupils with opportunities to be engaged in doing science through understanding both the argument structure of science and the importance of language in science.' (p.849). Previous work by Hand had shown this to be successful with older pupils (see Hand, 2008), and

positive effects were also found with the younger pupils. There was an average effect size of 0.3. As discussed in the previous section, the degree of fidelity with which the intervention was implemented impacted on its effectiveness. Notably, in the classrooms with a high degree of implementation fidelity, the effect on low-SES pupils (as measured by free or reduced price school meals) was greater than for other pupils: the gap appeared to reduce from a difference of 0.4 standard deviations to 0.2. In this case, pupil attainment was measured by the Iowa Test of Basic Skills. This looks promising, and the authors draw out three key points from their analysis to consider in future research. Firstly, the practice of argument in authentic embedded experiences was deemed crucial to pupil success. Secondly, this needs to be augmented by writing opportunities for pupils, using the science writing heuristic approach to promote a sense of science literacy. Finally, teachers need to be supported by an ongoing professional development programme, with a full three years being needed in this case to see improvements in pupils' language scores as well as their science scores.

Tong et al. (2014) conducted a study with five teachers and 94 11-year-old pupils (with seven teachers and 194 pupils as control) over a school year in two schools (and with two control schools) in Texas, where 85% of pupils received free or reduced price school lunches. The literacy-integrated science intervention comprised scripted lesson plans for daily oral and written activities with a science inquiry focus, comprehension exercises with scientific texts and the development of glossaries (to last 85 minutes every day). Pupil attainment was assessed using the state-wide standardised assessments and a district-development benchmark test for science and reading. The researchers reported that the treatment pupils were 10.28 times more likely to pass the tests than the control pupils, which was a significant positive effect. They attribute this success to the combination of the literacy intervention with the inquiry-based learning cycle, stating that 'inquiry-based approaches may not be sufficient enough for those pupils who are underachieving in both science and reading, because reading is critical to science learning' (p.2103). This sort of intervention clearly requires a strong commitment on the part of a school and its staff, not least because the professional development component requires two hours of staff time every fortnight, raising the question of how this might be embedded in a more sustainable way. The same research team carried out additional work developing and validating the rubric to be used to analyse pupils' science notebooks (Huerta et al. 2014). Here, their focus was on pupils with English as an Additional Language (EAL), and their rubric was based on second language acquisition theory. Although this sample was only 30 pupils, the process highlights the importance of careful instrument design to ensure the intended constructs are measured.

Cromley et al. (2013) worked with one teacher, who received two hours of professional development, and 61 pupils, aged 15, in a selective school for low-income pupils over two months. They measured parental education (as reported by pupils) as a measure for SES, with 53% or less of mothers having graduated from high, and 85% of fathers. Pupil attainment was measured by testing pupils' understanding of biology diagrams, with an effect size of 0.29 (literal comprehension) and 0.52 (inferential comprehension). The control group (consisting of the rest of the year group) had pre- to post-test effect sizes of 0.15 and 0.19 respectively. Although this is a narrow aspect of the science curriculum, it appears to be a promising approach and perhaps could form a small part of a larger intervention.

Three other studies examined literacy-based interventions using a range of methodologies: Hanrahan (1999) and Lyon (2013) used case studies, while Lee et al. (2006) carried out a pilot for a RCT, followed by a pre/post-test experimental design (2008). Hanrahan (1999) worked with a class of 24 12 year olds in a school situated in a low-SES area in Australia. Her intervention was to introduce journals for pupils to record their thoughts and questions about science, to which she gave individual written feedback. This class had particularly low levels of literacy compared to other classes in the school, but over the year this opportunity for pupils to 'talk science' (Lemke, 1990) in an affirmational and dialogic way appeared to lead to a classroom where the teacher spent more time on language and literacy in science. The class obtained comparable results to those of the other classes in the end-of-year-tests, despite earlier negative expectations. Clearly, this needs further research, but the simple intervention appeared to impact on teacher practice and pupil learning at multiple levels.

Lyon (2013) compiled 3 case studies of pre-service secondary science teachers looking at their use of language in assessment, with a focus on equity issues. Located in rural and urban California in schools with culturally, socially and linguistically diverse populations, the rich and detailed accounts give a clear picture of how these white, middle-class pre-service teachers became more aware of language issues in science and the barriers language can pose to learning science. The pre-service teachers learned to navigate the tensions between reducing the language demand of activities and providing additional scaffolding to enable pupils to access the activities. They also grappled with the issue of how to disentangle the assessment of language competence and of scientific understanding. Working from the premise that equitable science assessment can leverage learning opportunities, this study raises questions about initial teacher education and the extent to which learning how to teach and to assess language in science is part of every pre-service science teacher's experience.

Lee and her team have published extensively (e.g. 2006, 2008) about their research with culturally and linguistically diverse elementary school pupils, often working in schools with 80%+ low-SES (free or reduced price school meals). The 2 year intervention provided schools with teaching units aiming to promote inquiry-based learning and to support pupils' language and literacy needs, as many of the 1600 pupils also had English as an Additional Language. These were accompanied by 8 days of professional development for the 56 teachers over the 2 years. Effect sizes are reported for the different year groups, ranging from 0.79-2.86, but not for the low-SES pupils within each year group. Although all pupils appear to benefit from the intervention, the data suggest that the low-SES pupils benefit less than the high-SES pupils, i.e. the attainment gap between these two groups of pupils appears to widen rather than decrease. Naturally this was of concern to the researchers, but no solutions were presented in the publications that we found.

Summary of research about literacy in science

In summary, the research reviewed here on the use of literacy to promote science attainment for low-SES pupils suggests that this is also an area where interventions can be effective in improving the science attainment for low-SES pupils. As with the scientific reasoning and inquiry research, curriculum initiatives accompanied by professional development for teachers seem promising. The caveats are also similar: time is needed for the full effects to be seen, which raises questions about the costs and sustainability of the interventions, and how vulnerable they are to curriculum and structural changes in schools. An additional caveat is for the intervention to be tailored to the literacy needs of the pupils

concerned: many of these studies were conducted in the USA with large minority populations of Hispanic pupils. The pupil population in the UK is considerably different, as are the language needs of our low-SES pupils, and any intervention should be designed to take these into account. Again, the majority of research has focused on primary and lower secondary school pupils with promising results, but little has been done with older secondary school pupils. As well as interventions designed to support pupils' understanding and use of science-specific language, work on science-specific linguistic tools, such as diagrams, also appears to be promising.

c.   Studies using technology to support pupils' learning and engagement

As Table 6.1 shows, the search produced ten studies focused on the use of technology to support pupils' learning and engagement in science education, and these will be grouped by the type of technology used: laptops and ipads for pupils, and the use of technology to support teachers' professional development. The study by Marx et al. (2004) has already been discussed in an earlier section, but the age of this source makes it less relevant for a discussion about technology. These studies were all conducted in the USA.

Studies investigating the effect of giving pupils laptops or i-pads included Zucker & Hug (2008), Zheng et al. (2014) and Boyce et al. (2014). Zucker & Hug surveyed 311 pupils, aged 14-18 years, in a low-SES school (40% low income families) where every pupil had a laptop. They reported positive findings, but did not measure science attainment. Zheng et al. conducted a quasi-experimental study investigating the effect of using laptops and interactive science software over a year with 10 year old pupils at schools with high percentages of free/reduced price lunches (the percentages were not given), but did not find a significant positive effect. Boyce et al. (2014) gave i-pads to 55 pupils, aged 10, from two low-income schools on 2 nature hikes, and through interviews and monitoring the use of the i-pads, found that the pupils used these to photograph organisms and take notes.

Ye et al. (2015) and Blanchard et al. (2016) investigated the effect of using technology to support teacher professional development. Over a year Ye et al. (2015) worked in a quasi-experimental way with 73 teachers and their 14 year old pupils in schools with 26% - 73% pupils receiving free or reduced price lunches. This was done through the provision of a web-based application to support teachers' planning and subject knowledge development. Teachers' usage of the application was logged, and a pre- and post-survey of teachers showed a significant increased awareness of other teachers' practices in teaching Earth Science and in the frequency of using interactive resources in lesson planning and teaching. Pupil attainment was measured pre- and post- intervention with effect sizes reports of 0 to 0.71, but there was no control group. The variation appeared to be due to the variety of ways teachers used the technology and resources (high variety equating to higher effect sizes), and also due to variations in implementation across different school districts, reflecting earlier discussions. Blanchard et al. (2016) conducted a three year study with 20 teachers in two middle schools in rural, high-poverty areas, with the teachers receiving professional development to integrate technology into their teaching. Evaluation of teachers' practice did not seem to suggest much change, but pupils appeared to achieve slightly higher in the end of year assessments in mathematics and science. Together, these studies suggest tentative support for the use of technology to support teachers' professional development, but care needs to be taken with implementation.

Finally Plass et al. (2012) investigated the use of simulations to support learning in chemistry. The simulations covered key aspects of chemistry including kinetic theory, diffusion, phase changes and gas laws. The four year research project worked with 11 teachers and 718 pupils (227 control), aged 15-16, in a quasi-experimental design and found effects of up to 0.56 in pupil attainment, however, the SES of the pupils is not reported. This appears worth further investigation.

Summary of research on the use of technology to support pupil learning and engagement

The research on the use of technology to support pupil engagement and learning in science education is not extensive. It does not appear that providing pupils with laptops or other devices has a significant effect, but the literature reviewed does suggest that the use of technology to support teachers' professional development and the use of simulations has the potential to benefit pupils.

d. Studies aiming to develop pupils' metacognitive skills

The study of metacognition with low-SES pupils in science education is a small field of research (see Table 17). In Chapter 4 we discussed two successful metacognition interventions (Zohar & Peled, 2008; White & Frederiksen, 1998), both of which benefitted low attaining pupils more than high attaining pupils. We will not describe these two studies any further in this chapter which concentrates on measuring the effects of teaching programmes on low SES children, though it is worth noting here that the majority of the participants in White and Frederiksen's study came from minority backgrounds.

Of the relevant articles that our search did produce, only three reported effect sizes. One of these, Oliver et al. (2012), has also been discussed in the earlier section on scientific reasoning and inquiry skills, as the principles underpinning CASE include both reasoning patterns and metacognition (effect size of 0.47 for cognitive gain). However, it is pertinent to add here that Oliver et al. also evaluated pupils' perceptions as well as attainment, and a number of pupils were able to articulate how the CASE lessons had enabled them to think more and to think differently about the science investigations they were conducting. 58% of the pupils also identified this part of the lesson as the most difficult aspect.

Fouche's USA-based doctoral studies (2013) used a non-equivalent control group design with stratification and multiple control groups to examine the effect of encouraging pupils to use metacognitive and self-regulatory strategies. 215 pupils (88 control), aged 12-15 years, were given their own performance data, and were shown how to evaluate these data and then use them to plan and inform their own learning. There was an emphasis on enabling pupils to take more control of their learning. The setting was unusual in that this was a residential school for low-SES pupils who were 'exclusively from families in extreme poverty and social need from select communities across the country' (p.59). The intervention lasted 6 months and the effect size on pupils' attainment in the school's standard physics tests was 0.42. This looks promising, but the author rightly recognises that more research is needed, with a larger sample of non-residential pupils over a longer period of time.

The *Responsive Classroom* Approach was investigated by Griggs et al. (2013) using a RCT with 5th grade pupils [age 10] in a 'large ethnically and socioeconomically diverse school district located in a mid-Atlantic state, USA' (p.364), although the actual levels of low-SES are not reported. 797 pupils from 12 schools took part in the year-long intervention, with

764 pupils from another 8 schools as the control sample. The intervention uses a series of principles, recognising that '*how children learn is as important as what they learn*' (p.362), and practices '*to help teachers create safe, supportive classroom climates conducive to academic learning*' (p.362). Pupils' level of anxiety about mathematics and science were measured, as were their self-efficacy beliefs for both subjects, with an effect size of 0.24. Pupils' attainment in science was not measured. Although the authors identify anxiety as a barrier to learning, and low-SES pupils seem to have higher levels of anxiety and lower levels of self-efficacy, the evidence presented is not conclusive about what effect reducing pupils' anxiety and increasing their self-efficacy beliefs will have on pupils' attainment.

Baines et al. (2007) conducted a quasi-experimental design study designed to improve group work in primary school science lessons. 265 pupils aged 9 (486 control) and 295 pupils aged 14 (541 control) took part in the intervention from 12 schools in London, UK, with an average of 35.8% FSM. The intervention lasted 4 months, with 7 half day meetings as professional development for the teachers involved. Pupil attainment in science was measured via a variety of tests, and appeared to show a positive impact with the intervention pupils scoring higher than the control pupils by between 0.23 and 0.5 standard deviations. However, when the % of FSM increased, there was a slight decrease in attainment by ~0.1 standard deviations. The authors still felt that all categories of pupil benefitted from the intervention, including pupils with low prior attainment.

Summary of research about metacognitive skills

To summarise, although Chapter 4 identifies pupils' metacognitive skills as a possible mediator of their attainment in science, the number of studies which look at interventions with low-SES pupils studying science is disappointingly low. Nevertheless, the evidence discussed here suggests that interventions designed to develop pupils' group work skills, and to teach them to evaluate and make use of their own assessment data are beneficial to their learning of science. Similarly, the CASE programme has a positive contribution here to help pupils to think about their science learning as well as to develop pupils' scientific reasoning skills. Again, this research tends to focus on pupils in upper primary and lower secondary schools.

e. Studies from a socio-cultural perspective

This group of studies was by far the largest of those included in the review, though overwhelmingly these were qualitative in their methodological approach. In addition, these studies tended to be concerned with increasing participation in science education from the point when it is no longer compulsory, rather than improving science attainment. As one might expect from a socio-cultural perspective, the researchers in each case were investigating ways to break down the barriers to participating in science education which are perceived by low-SES pupils, and tended to approach this in three ways:

  i. Bringing pupils into a science 'place' e.g. university laboratories (Amos & Reiss, 2012) or a science museum (Dawson, 2014b)
 ii. Bringing scientists or extra-curricular science activities into schools (e.g. Gartland, 2015; Mosatche et al., 2013; Schneider et al., 2012)
iii. Developing teachers' understanding of pupils' perspectives (e.g. Yerrick & Johnson, 2011; Tobin, 2005; Johnson & Fargo, 2014)

Table 6.2 gives an overview of key features of these studies, whose main purposes were to improve pupils' engagement and motivation in science lessons by making it seem more relevant and accessible to them, thereby hopefully improving their attainment in science and encouraging them to participate further in science education:

Table 6. 2 Overview of key features for studies using a socio-cultural perspective.

	Bringing pupils into a science 'place'	Bringing science & scientists into school	Developing teachers' understanding of pupils' perspectives
Number of studies	14	12	22
Number UK-based	2	3	0
Number with sample size greater than 30	9	10	9
Number where intervention is longer than 1 month	7	11	10

The number of studies in the first and second groups was roughly equal, but the third group was double in size, with this being an interesting point in itself. It could be argued that although taking pupils out of their everyday environment and putting them into a science 'place', or bringing science and scientists into schools can be effective at breaking down barriers, the educational research community seems to think that working with teachers and equipping them to break down these barriers in their classrooms on a daily basis has the greatest potential to impact on pupils' attainment and participation in science. The first group could also be sub-divided into approaches initiated by teachers and schools (12 studies), and approaches initiated by the participants or their families (two studies). However, it is not clear whether the locus of control is important. The studies cited in the list above will be discussed as exemplars of these three types of initiative, before discussing the wider literature.

Bringing pupils into a science 'place'

Two examples will be discussed here: the study by Amos and Reiss (2012) of providing residential fieldwork for science, an example of where the locus of control is with the school and teachers, and a contrasting example of a study by Dawson (2014b) into the use of informal science education institutions by low-income minority ethnic groups, in which the locus of control is with individuals and families.

Amos and Reiss (2012) examined the benefits of a five year initiative to provide residential science fieldwork for 11-14 year old pupils in London schools, with 33 000 pupils from 849 schools taking part overall as part of the London Challenge Programme. The evaluation used pre- and post- course questionnaires, interviews and observational data from 2706 pupils (pupil interviews were only post-course and were focus groups), 70 teachers and 869 parents/carers from 46 schools which participated, (no control). During the first three years, the focus was on schools with predominantly low-SES pupil populations, but the programme was opened up during the last two years – 35 of schools had pupil cohorts with SES over 16 (using the Fischer Family Trust scale of 1-27, so relatively high levels of deprivation). The findings revealed gains across a range of areas, including social, affective, physical, and cognitive: pupils developed their cooperative learning skills and their self-efficacy beliefs

through the physical challenges, and the three-five day courses 'immersed pupils in the development of practical inquiry skills, over several days, allowing the revisiting of ideas and challenges, and occasionally re-running investigations' (p.496), thus providing an opportunity to make science learning more authentic which can often be unavailable to science teachers in schools with an extensive curriculum to cover. The researchers were positive about the benefits of residential science fieldwork for these low-SES pupils, arguing that they provided '*much-needed variety, stimulus and reality to educational experience and well-being*' (p.507).

The residential fieldwork was clearly organised through and by schools, although parents and carers evaluated it positively, as did the pupils. By contrast, Dawson's (2014b) qualitative study over a year with 58 participants (seven under 16) from central London community groups on facilitated visits to three informal science education institutions i.e. science museums and science centres, found that the institutions' 'practices were grounded in expectations about visitors' scientific knowledge, language skills, and finances in ways that were problematic for participants and excluded them from science learning opportunities' (p.981). Although not exclusively focused on school age pupils, this study provides interesting insights into why low-SES pupils and their families might perceive science – both learning about it and practising it – as something from which they are excluded. As participation in science outside school can be considered part of a pupil's '*science capital*', which has been linked to participation in science post-16 (Archer et al., 2015), this study provides some useful suggestions about how informal science education institutions could attempt to work more with low-SES pupils and their communities. The importance of encouraging positive parental attitudes towards science was highlighted by a review of the 2006 PISA data (Perera, 2014), as these have been shown to have a positive and statistically significant effect on science achievement, with low-SES pupils benefiting as much as high-SES pupils.

Bringing science and scientists into school

Examples in this category include two papers reviewing extra-curricular STEM-focused interventions in schools in urban areas in the USA (Mosatche et al., 2013; Schneider et al., 2012) and an investigation into the impact of STEM pupil ambassador outreach programmes in the UK (Gartland, 2015). Mosatche et al. (2013) examined three STEM programmes for girls aged 10-18 with pre- and post- surveys of participants (237, 1234 and 121), comparison groups (details not given), and interviews with parents and teachers (again, details not given). Two of the programmes lasted from one-six years, and the other from one day to one year. They suggest that the following are needed for such interventions to be successful: the science (or engineering/technology/mathematics) programme needs to be engaging and relevant, allowing participants to explore concepts in depth over time. Part of this should involve the opportunities to explore careers, exposing participants to role models and providing opportunities for field trips as well, with these findings supporting the points made by Amos & Reiss (2012). Echoing points made earlier about professional development, Mosatche et al. emphasise the need for the adults leading the programme to have both a good knowledge and understanding of the science concerned, *and* a good understanding of how to work with girls of this age group and background – how to maintain the '*fun factor*' (p.21)!

Similarly, Schneider et al. (2012) report on the *College Ambition Program* (CAP) at Michigan State University, which, motivated by the 20% difference between low-SES and high-SES

pupils at college, aims to encourage high school pupils to progress onto college and to study STEM subjects at college. Working with eight schools and 53 STEM undergraduates to act as mentors, the program has 4 strands: '*mentoring and tutoring, course counselling and advising, college visits and financial aid guidance*' (p.63), with the latter being provided by a specially commissioned publication, '*Ten steps to college*' by L. Beasley-Wojick. Preliminary results show that pupils from these CAP schools are more likely to attend college than pupils from similar schools (data and methodology not given), with stronger effects for the urban schools compared to the rural schools. They also report stronger self-efficacy beliefs, and requests for the programme to be extended to other schools. These tentative findings seem to reflect those reported by Dawson (2014b), namely that the barriers to science education perceived by low-SES pupils are multi-faceted, and interventions should be designed to address multiple issues. This is further supported by the correlational study with 3223 pupils by Gottfried and Williams (2013) which showed little association between STEM club participation and STEM outcomes for low-SES pupils, suggesting that STEM clubs on their own are not sufficient to close the gap in STEM achievement between low-SES pupils and the cohort as a whole.

The final example to be discussed in this section is the ethnographic study by Gartland (2015) which explores the notion of pupil ambassadors as role models in more detail. Working with 32 STEM undergraduate ambassadors at two London universities over two years, and 112 pupils, aged 12-16, predominantly from 'deprived' south-east London boroughs, the findings from observations and interviews suggest that some of the assumptions about the benefits of role-models are flawed if the interaction is too formal. However, more informal, subject-specific interactions, based around experiential problem-based learning, where both the role model and the school pupil were working together as learners, enabled the school pupils to develop a '*science identity*' (linked to the notion of '*science* capital' referenced earlier (Archer et al., 2015)), and to see themselves as possibly studying STEM subjects at university. Again, this reflects the findings in Section a) about the benefits of teaching and learning which is designed to develop pupils' scientific reasoning and inquiry skills. This is also further supported by Curtis et al. (2012) who evaluated an Australian peer-mentoring programme with 46 pupils aged 14-16 (over two years) from one rural school and one low-SES school and found pupils who consistently received high or moderate levels of mentoring reported higher chances of attending university (65%) compared to those who received little or no mentoring (49%). This was not subject-specific, and had therefore been excluded from the list of studies in EPPI, but the long-term nature of this programme and the principled nature of the mentoring echo the findings of Gartland.

Developing teachers' understandings of pupils' perspectives
As shown in Table 6.2, this group of studies is the largest under this socio-cultural perspectives heading, and again, the USA dominates the field of study, though the one ethnographic example discussed here is from Australia with a teacher who shared a similar social and cultural history to her pupils, and the other from the USA with a teacher from a very different background to his pupils. A final example from the USA draws together culturally relevant pedagogy with inquiry-based learning (Johnson & Fargo, 2014).

Yerrick and Johnson (2011) spent two years constructing a rich case study from classroom videos and artefacts from lessons with one teacher and 32 predominantly black pupils aged 14-16 (over the course of the study) in a rural school in the USA. The class was following an earth science course while the teacher endeavoured to develop classroom discourse

through employment of various teaching strategies and through developing a better understanding of his pupils. In particular, the teacher found that his knowledge 'shifted', in his understanding of the 'disparity of the pupils' discourse from the syntactic structures of the scientific discipline, his pedagogical content knowledge which grew with pupils' reports of successful teaching events, and his knowledge of the pupils' educational context or milieu' (p.925). This ethnographic approach showed how the development of the teacher's knowledge allowed the growth of a classroom culture where pupils were asking scientific questions and developing scientific explanations: they felt connected to the curriculum and empowered to contribute to lessons. The teacher also became an advocate for 'his' pupils in school, negotiating on their behalf when relationships with school leaders broke down. Although the SES levels of the pupils are not explicitly stated, they are presented as a stark contrast to the white, middle-class, middle-aged male teacher and are clearly a group of low-attaining pupils with negative involvement with the police and high incidences of teenage pregnancy. Crucially, this study presents an account of how teachers can successfully learn to teach a group of pupils very different from themselves, and calls for initial science teacher education to include the development of 'knowledge of the pupils that will determine appropriate pedagogy for an effective and culturally responsive classroom' (p.937).

In contrast, Tobin (2005) presents an account of a teacher who was already recognised as successful in teaching 'working class, ethnically diverse pupils', and who 'had a strong orientation toward equity issues, and perceived herself as someone who had escaped from the bonds of working class life to become a university graduate and teacher' (p.580). This mixed methods study examined the teaching and learning of a chemistry topic over a five week period with one class of 31 pupils, aged 15-16 years, in a suburban high school in Australia, where the majority of families were working-class. It was found that the teacher encouraged participation in science lessons by the way she presented the purposes of science education as being relevant to their lives outside the classroom, ensuring that the classroom was a safe space to offer contributions, whether correct or not, promoting collaborative group working and clear structures for lesson participation. She also emphasised the importance of reading in science, explicitly teaching literacy skills, which reflects the points made in the previous section on literacy. The focus on active involvement in lessons and successful completion of tasks led to enjoyment and active participation in science lessons, with pupils learning knowledge and developing skills, which Tobin contrasts to the typical experiences of working class pupils. Clearly, not every class of low-SES pupils can be taught by a teacher from their own background, but this study offers some useful suggestions of how to encourage culturally adaptive teaching and learning of science.

The theme of culturally relevant pedagogy is continued in the case study by Johnson and Fargo (2014) in an urban elementary school in the USA. 10 teachers participated in a professional development programme over two years with two weeks of summer workshops, eight days workshops and 20 monthly support sessions, to plan and teach inquiry-based learning using culturally relevant pedagogy. This is defined as 'pedagogy which combines pupil knowledge, experiences, and cultures to enable pupils to be successful academically, exhibit cultural competence and become socio-politically critical' (p.847. Eighty-eight percent of the 52 pupils (51 control), aged 9-11 years over the course of the study, qualified for free or reduced price lunch (99% control). In contrast to the other

studies discussed in this section, the researchers did report pupil attainment in science, using the state standard assessments, which saw the treatment group achieve 67% in the assessment at the end of the two years (control achieved 29%). The researchers particularly noted that for Hispanic pupils, the treatment group achieved 85% compared with 25% for the control (Hispanic pupils were 62% of the pupil cohort in the treatment school, 74% in the control school). Whilst this study may seem more relevant to the section on inquiry and reasoning, it is discussed here to support the arguments already made about the importance of enabling teachers to present the science curriculum as culturally relevant to their pupils, and how teachers from backgrounds different to their pupils may need extra support to achieve this.

The discussion now moves on to wider studies, including review papers. One is the meta-analysis by McLaughlin (2014) which examined urban science education studies published between 2000 and 2013, with her list of 68 papers having a high degree of overlap with our, longer, list. She concluded that, although there have been many changes to educational structures and organisation, urban pupils are still 'oppressed' by systems that prioritise control and order over effective science education. She also identified potential ways forward: the need for teacher professional development in order to allow *pupils "to make relevant connections between cultural and linguistic funds of knowledge with the scientific discourse of their classroom" (p.905), and to move from "a pedagogy of poverty that focuses on low level didactic teaching strategies that detract from meaningful instruction and diminish the quality of pupils' learning"* (p.919). This latter point reflected the findings of Slavin et al. (2014) as discussed earlier.

More positive is the review of research on informal science education by Dawson (2014a) of nearly 150 articles, which she uses to develop a framework for access, equity and inclusion in informal science education, around infrastructure, literacy and community acceptance. Drawing parallels with schools, this review supports the suggestions made earlier that an emphasis on supporting the development of literacy in science, and working with teachers to develop their understanding and acceptance of different groups of pupils in schools would both be productive avenues for future research.

The final review paper is by Kaur (2012), which examined eight articles published in the journal *Teaching and Teacher Education* between 1990 and 2010 on issues of equity and social justice, with these being a subset of the list of over 130 articles. Like the research discussed so far, Kaur found that much research originated in the USA, and was often multi-faceted in its approach. He also found that much of the research was trying to investigate and develop a better understanding of these complex teaching and learning communities, rather than implement and evaluate particular interventions. Given the emphasis in the discussion so far on the importance of teacher professional development, which can also include initial teacher education, this provides some interesting insights into the field and reflects the rest of the literature discussed so far.

Summary of research from a socio-cultural perspective

Studies using the socio-cultural perspective to investigate the teaching and learning of science education for low-SES pupils dominate the research, and examples have been discussed under three headings: bringing pupils into a 'science' place, bringing science and scientists into school contexts, and developing teachers' understandings of pupils' perspectives. The studies discussed have provided evidence mostly of a qualitative nature,

which supports residential fieldwork and the use of informal science education institutions, whilst recognising that these need to be set up and used in carefully structured ways. Support is also given to the benefits of after-school activities, such as STEM clubs, and peer-mentoring, with the caveats that these also need detailed planning of professional development for the leaders involved. Finally, studies with teachers suggest that teachers can be supported to develop more culturally relevant pedagogy, though appropriate strategies for UK schools may differ greatly from those found to be successful in the USA and Australia.

Two important points should be made here: first that successful interventions appear to be designed to address more than one aim, reflecting the discussion in previous sections, and second, in contrast to earlier discussions, that many of these socio-cultural interventions are with older school pupils, often aged 14 and older.

Features of successful interventions and implications for future studies

At this point it seems profitable to reflect on what makes a successful intervention, given that some of the interventions reviewed appear to have had significant positive effects on outcomes for low-SES pupils, whilst others have had little or negative effects. Some of these features are well defined and easily controllable; others vary according to the local context and are more difficult to control.

The first feature has been discussed already throughout this chapter, namely the need for interventions to include professional development for the teachers and staff who will be using the pedagogical resources and strategies which are the focus of the intervention. Much has been written about teacher professional development (e.g. Corrigan et al., 2010), which will not be repeated here, but common across the more effective studies reviewed here has been the ongoing nature of this professional development throughout the length of the intervention. In addition, many studies have incorporated both teacher workshops (out of lessons) and coaching (in lessons), in order to develop teachers' understanding of the intervention and to support them in its implementation – allowing them to interpret it for their context, yet maintaining a high fidelity. This all has to be accomplished with a manageable time commitment for a profession which already has long working hours – a challenge indeed!

The length of the intervention is the next key feature: the majority of these studies had a duration longer than one month, many lasted for a school year, and a number lasted three-five years, with effects only being seen in the second or third years. For some of the shorter interventions, the researchers felt that the time had been insufficient to see an effect, or questioned the longevity of the effects observed.

The question of how to integrate the intervention into pupils' and teachers' practice and experiences is also a key feature: some of these studies required daily sessions, others weekly, but the effective ones were those used regularly. An interesting question is to what extent these skills and resources continue to be used once the intervention has ended.

The teaching and learning of science is highly complex, and low-SES pupils may need support in more than one area. Some of the successful interventions aimed to address more than one issue at a time, tailoring these according to the learning needs of the pupils concerned, e.g. for some pupils it may be appropriate to focus on developing their scientific reasoning skills *and* their literacy skills in order to articulate their reasoning.

Finally, it is clear from all these interventions that support from school leadership is crucial to their success, with this support being in the form of finance, time and encouragement. Unfortunately, some of these interventions were undermined by changes to school structures, curriculum and assessment.

Although not included in the EPPI database, the review by Bell (2014) of successful science interventions should also be mentioned here. Commissioned by the Wellcome Trust and building on the report by Falk et al. (2012), he interviewed 45 science teachers and educators about 57 UK-based interventions. Not all of these interventions had been evaluated in a robust way, though nearly all were perceived to have had a positive impact on pupils' experiences of science education. He concluded that there were three key elements necessary for successful interventions: 'underpinning principles of the intervention, the expertise of the personnel and the context of the intervention' (p.3), and made a number of recommendations about these which are listed in Appendix 6.3.

Having considered these key features of successful interventions, the question arises as to what implications this review has for future work? It is clear from the literature reviewed that more research would be productive in the areas considered: what interventions would support the development of pupils' scientific reasoning and inquiry skills, their literacy skills and their metacognitive skills; what role technology should play in this work; and what contributions can socio-cultural perspectives offer to support the attainment and progression of low-SES pupils in science education. All of these clearly have the potential to have a positive impact on low-SES pupils' learning of and participation in science, though with a number of caveats that have been discussed already and are summarised below.

Table 6.1 strongly suggests that particular interventions can be effective with certain age groups of pupils. Research in primary schools and lower secondary has the potential to impact positively and significantly on the development of pupils' skills. Given that, if there are gaps in these areas, they are likely to be smaller when pupils are younger and bigger when pupils are older, intervening earlier in order to support pupils and 'close' the smaller gap would appear more logical, and somewhat easier, than when the gap is bigger. However, the socio-cultural work benefits older pupils, again if pupils are struggling to see themselves as 'scientists'. Working with them throughout the adolescence period while they attend secondary school should have the potential to make a positive impact.

It is also clear that interventions need to have clear aims and careful research design, such as the choice of control group (if there is one), and well defined outcome measures which match these aims and appropriate instruments to measure them e.g. attainment and self-efficacy. Importantly, interventions should be tailored to their context e.g. in what ways do *these* pupils need to develop their scientific reasoning skills? This suggests the need for careful and extensive baseline data collection.

Finally there is the question of sustainability: if these interventions are found to be effective, can these pedagogical strategies and resources be embedded in teachers' and schools' practice so that this becomes ongoing? How might this link with the research agenda in schools and the use of evidence-based practice, so that the benefits continue even when the research funding is ended?

Conclusion

In conclusion, this review of the literature has revealed a number of promising educational approaches to support the attainment and progression of low-SES pupils in science education. As discussed in the preceding section, there are several implications for future research, and clearly new UK-based studies in the areas described above would have the potential to be beneficial to low-SES pupils. There are also implications for teacher educators: how might pre-service science teachers be educated to develop pupils' reasoning and inquiry skills, their literacy skills and their metacognitive skills? How could pre-service science teachers be supported in their use of technology and enabled to develop a socio-cultural informed pedagogy? These questions are also relevant for the wider science teaching profession, and it is hoped that the findings from this literature review will encourage and foster ongoing conversations about how teachers and schools can work together, with resulting changes to practice, in order to increase the attainment and participation of low-SES pupils in science education.

Table of acronyms

ALSPAC	Avon Longitudinal Study of Parents & Children
APS	Average Points Score
AT	Attainment Target
BTEC	Business & Technician Education Council. A National Qualification equivalent in vocational areas such as Nursery Nursing, Business Studies and Art and Design. There are considerable practical elements to the courses with work placements offered
C AP	College Ambition Program (USA)
CASE	Cognitive Acceleration through Science Education programme
CVS	Control of variables strategy
DFE	Department For Education
EBACC	English Baccalaureate
EEF	Educational Endowment Foundation
ES	Effect Size (See Appendix 2.2)
EVER6	A pupil who has been entitled to a Free School Meal (FSM) at anytime during the past six years
EYFSP	Early Years Foundation Stage Profile
FSM	A pupil entitled to Free School Meals
FTE	Full-Time Education
GCSE	General Certificate in Education
GDP	Gross Domestic Product
IAE	International Association for the Evaluation of Educational Achievement
IDACI	Income Deprivation Affecting Children Index
KS	Key Stage
LA	Local Authority
NAEP	National Assessment of Educational Progress (USA)
NARA	Neale Analysis of Reading Ability – a reading comprehension measure
NASS	National Asylum Support Service
NC	National Curriculum
NPD	National Pupil Database (UK)
OECD	Organisation for Economic Co-operation and Development
OR	Odds Ratio (See Appendix 2.2)
PISA	Programme for International Student Assessment
PRU	Pupil Referral Unit: A specialist establishment designed for the education of young people excluded from mainstream school
ROSE	The Relevance of Science Education project (USA)
SD	Standard Deviation
SEN	Special Educational Needs
SES	Socio-economic Status
STEM	Science, Technology, Engineering & Mathematics
TIMSS	**Trends in International Mathematics and Science Study**
YR	Reception Year
WISC	Wechsler Intelligence Scale for Children

See also Cambridge Handbook of Educational Abbreviations and Terms (CHEAT) 2013. See:
https://www.educ.cam.ac.uk/people/staff/hickman/CHEAT_RichardHickman_6th_ed%20_2013.pdf

References

Adey P. S. & Shayer, M. (1990). Accelerating the development of formal thinking in middle and high school students. *Journal of Research in Science Teaching, 27* (3), 267-285.

Ainley, M., & Ainley, J. (2011). Student engagement with science in early adolescence: The contribution of enjoyment to students' continuing interest in learning about science. *Contemporary Educational Psychology, 36*(1), 4-12.

Alexander, K. L., & Cook, M. A. (1982). Curricula and coursework: A surprise ending to a familiar story. *American Sociological Review*, 626-640.

Amos, R. & Reiss, M., (2012). The Benefits of Residential Fieldwork for School Science: Insights from a five-year initiative for inner-city students in the UK, *International Journal of Science Education*, 34(4), 485-511.

Anderson, J. O., Lin, H.-S., Treagust, D. F., Ross, S. P., & Yore, L. D. (2007). Using large-scale assessment datasets for research in science and mathematics education: Programme for International Student Assessment (PISA). *International Journal of Science and Mathematics Education, 5*(4), 591-614.

Aram, D., & Levin, I. (2001). Mother–child joint writing in low SES: Sociocultural factors, maternal mediation, and emergent literacy. *Cognitive Development, 16*(3), 831-852.

Ardila, A., Rosselli, M., Matute, E. & Guajardo, S. (2005) The Influence of the Parents' Educational Level on the Development of Executive Functions, *Developmental Neuropsychology*, 28:1, 539-560, DOI: 10.1207/s15326942dn2801_5

Areepattamannil, Shaljan, & Kaur, Berinderjeet. (2013). Factors predicting science attainment of immigrant and non-immigrant students: a multilevel analysis. *International Journal of Science and Mathematics Education, 11*, 1183-1207. doi: 10.1007/s10763-012-9369-5

Baddeley, A. (2000). The episodic buffer: a new component of working memory? *Trends in cognitive sciences, 4*(11), 417-423.

Baines, E., Blatchford, P. & Chowne, A., (2007). Improving the effectiveness of collaborative group work in primary schools: effects on science attainment, *British Educational Research Journal*, 33(5), 663-680.

Baird, J. R. (1986). Improving learning through enhanced metacognition: a classroom study. *European Journal of Science Education, 8*(3), 263-282.

Baker, D. P., Goesling, B., & LeTendre, G. K. (2002). Socioeconomic Status, School Quality, and National Economic Development: A Cross-National Analysis of the "Heyneman-Loxley Effect" on Mathematics and Science Attainment. *Comparative education review, 46*(3), 291-312. doi: 0010/4086/2002/4603-0002

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*(6), 1173.

Beese, J., & Liang, X. (2010). Do resources matter? PISA science attainment comparisons between students in the United States, Canada and Finland. *Improving Schools, 13*(3), 266-279.

Bell, D., (2014). The perceived success of interventions in science education, London, UK: Wellcome Trust.

Bernstein, B. 1975. Class, Codes and Control. Vol. 3. London: Routledge and Kegan Paul.

Bitner, B. L. (1991). Formal operational reasoning modes: Predictors of critical thinking abilities and grades assigned by teachers in science and mathematics for students in grades nine through twelve. *Journal of Research in Science Teaching, 28*, 265–274.

Blanchard, M. R., LePrevost, C. E., Dell Tolin, A. & Gutierrez, K. S., (2016). Investigating Technology-Enhanced Teacher Professional Development in Rural, High-Poverty Middle Schools, *Educational Researcher*, 45(3), 207-220.

Bornstein, M. H., Hahn, C-S., Suwalsky, J. T. D., & Haynes, O. M. (2012). Socioeconomic status, parenting, and child development: The Hollingshead four-factor index of social status and the socioeconomic index of occupations. In M. H. Bornstein & R. H. Bradley (Eds.), *Socioeconomic status, parenting, and child development* (pp. 29-82). Mahwah (NJ): Lawrence Erlbaum Associates.

Boyce, C. J., Mishra, C., Halverson, K. L., & Thomas, A. K., (2014). Getting Students Outside: Using Technology as a Way to Stimulate Engagement, *Journal of Science Education Technology*, 23, 815-826.

Boyd, A., Golding, J., Macleod, J., Lawlor, D. A, Fraser, A., Henderson, J., . . . Smith, G. D. (2012). Cohort Profile: The 'Children of the 90s'—the index offspring of the Avon Longitudinal Study of Parents and Children. *International Journal of Epidemiology, Apr. 16*, 1-17.

Bradley, L., & Bryant, P. E. (1983). Categorising sounds and learning to read -a causal connection. *Nature, 301*, 419-521.

Bradley, R. H. and R. F. Corwyn (2002). "Socioeconomic status and child development." *Annual review of psychology*, 53(1), 371-399.

Brown, A. L. (1980). Metacognitive development and reading. In R. J. Spiro, B. C. Bruce and W. F. Brewer (eds), *Theoretical Issues in Reading Comprehension: Perspectives from Cognitive Psychology, Linguistics, Artificial Intelligence, and Education* (pp 453-481). Hillsdale, NJ: Erlbaum.

Brown, A. L., Campione, J. C., & Day, J. D. 1981, Learning to learn: on training students to learn from texts. *Educational Researcher, 10*, 14-21.

Brown, A., Bransford, J., Ferrara, R., & Campione, J. (1983). Learning, remembering, and under-standing. In J. H. Flavell & E. M. Markman (Eds.), Handbook of child psychology: Vol. 3. Cognitive development (4th ed., pp. 77-166). New York: Wiley.

Brown, P. L., & Concannon, J. P. (2016). Students' perceptions of vocabulary knowledge and learning in a middle school science classroom. *International Journal of Science Education, 38*(3), 391-408.

Bryant, P., Nunes, T., & Barros, R. (2014). The connection between children's learning about grapho-phonic units and morphemic spelling rules and their attainments at school. *British Journal of Educational Psychology, 84*(2), 211-225.

Bryant, P., Nunes, T., Hillier, J., Gilroy, C., & Barros, R. (2015). The importance of being able to deal with variables in learning science. *International Journal of Science and Mathematics Education, 13*(1), 145-163.

Bullock, M. & Ziegler, A. (1999). Scientific reasoning: Developmental and individual differences. In F. E. Weinert & W. Schneider (Eds.), *Individual development from 3 to 12 (*pp. 38–54). Cambridge: Cambridge University Press.

Bullock, M., Sodian, B. & Koerber, S. (2009). Doing experiments and understanding science: Development of scientific reasoning from childhood to adulthood. In W.Schneider & M. Bullock (Eds*.), Human development from early childhood to early adulthood. Findings from the Munich Longitudinal Study* (pp. pp. 173–197). Mahwah,NJ: Lawrence Erlbaum.

Bybee, R., & McCrae, B. (2011). Scientific literacy and student attitudes: Perspectives from PISA 2006 science. *International Journal of Science Education, 33*(1), 7-26.

Byrnes, J. P., & Miller, D. C. (2007). The relative importance of predictors of math and science attainment: An opportunity–propensity analysis. *Contemporary Educational Psychology, 32,* 599-629.

Byrnes, J. P., & Miller-Coto, D. (2016). The growth of mathematics and reading skills in segregated and diverse schools: An opportunity-propensity analysis of a national database. *Contemporary Educational Psychology, 46,* 34–51.

Cano, F., García, A., Berbén, A. B. G., & Justicia, F. (2014). Science Learning: A path analysis of its links with reading comprehension, question-asking in class and science attainment, *International Journal of Science Education, 36*(10), 1710-1732.

Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology, 54*(1), 1.

Cattle, J., & Howie, D. (2008). An Evaluation of a School Programme for the Development of Thinking Skills through the CASE@KS1 Approach, *International Journal of Science Education, 30*(2), 185-202.

Chen, Z. & Klahr, D. (1999). All other things being equal: Acquisition and transfer of the control of variables strategy. *Child Development, 70*, 1098–1120.

Cohen, J. (1988). Statistical power analysis for the behavioural sciences (2nd Edition). Hillsdale, NJ: Erlbaum.

Comber, L. C., & Keeves, J. (1973). Science achievement in nineteen countries. New York: Wiley.

Corrigan, D., Dillon, J. & Gunstone, R., (2010). *The Professional Knowledge Base of Science Teaching*, Dordrecht, Netherlands: Springer.

Cromley, J. G. (2009). Reading Attainment and Science Proficiency: International Comparisons From the Programme on International Student Assessment. *Reading Psychology, 30*(2), 89-118.

Cromley, J. G., Perez, T. C., Fitzhugh, S. L., Newcombe, N. S., Wills, T. W. & Tanaka, J. C., (2013). Improving Students' Diagram Comprehension with Classroom Instruction, *The Journal of Experimental Education*, 81(4,) 511-537.

Crump, C., Ned, J. & Winkleby, M.A., (2015). The Stanford Medical Youth Science Program: educational and science-related outcomes, *Advances in Health Science Education*, 20, 457-466.

Curtis, D., Drummond, A., Halsey, .J & Lawson, M., (2012). *Peer mentoring of students in rural and low socioeconomic status schools: increasing aspirations for higher education*, Adelaide, Australia: National Vocational Education and Training Research Program.

Dawson, E. (2014). "Not Designed for Us": How Science Museums and Science Centers Socially Exclude Low-income, Minority Ethnic Groups. *Science Education, 98*(6), 981-1008.

Dawson, E., (2014a). Equity in informal science education: developing an access and equity framework for science museums and science centres, *Studies in Science Education*, 50(2), 209-247.

Dawson, E., (2014b). "Not Designed for Us": How Science Museums and Science Centers Socially Exclude Low-Income, Minority Ethnic Groups, *Science Education*, 98(6), 981-1008.

DeBacker, T., & Nelson, R. (2000). Motivation to Learn Science: Differences Related to Gender, Class Type, and Ability. *The Journal of Educational Research, 93*(4), 245-254.

DFE (2015). Level 2 and 3 attainment by young people in England measured using matched administrative data: Attainment by age 19 in 2014.

DFE (2016). Revised GCSE and equivalent results in England, 2014 to 2015 (SFR 01/2016).

Diaconu, D. V., Radigan, J., Suskavcevic, M., & Nichol, C., (2012). A Multi-Year Study of the Impact of the Rice Model Teacher Professional Development on Elementary Science Teachers, *International Journal of Science Education*, 34(6), 855-877.

Dilworth-Bart, J. E. (2012). Does executive function mediate SES and home quality associations with academic readiness? *Early Childhood Research Quarterly, 27*(3): 416-425.

Duit, R., & Treagust, D. F. (2003). Conceptual change: A powerful framework for improving science teaching and learning. *International Journal of Science Education, 25*(6), 671-688.

Donnelly, D.F, Linn, M.C. & Ludvigsen, S, (2014). Impacts and Characteristics of Computer-Based Science Inquiry Learning Environments for Precollege Students, *Review of Educational Research, 84(4), 572-608.*

Doppelt, Y., Schunn, C. D., Silk, E., Mehalik, M. M., Reynolds, B., & Ward, E., (2009). Evaluating the impact of a facilitated learning community approach to professional development on teacher practice and student achievement, *Research in Science & Technological Education*, 27(3), 339-354.

Duit, R., & Treagust, D. F. (2003). Conceptual change: A powerful framework for improving science teaching and learning. *International Journal of Science Education, 25*(6), 671-688.

Duncan, G. J., & Magnusson, K. A. (2012). Off with Hollingshead: Socioeconomic resources, parenting, and child development. In M. H. Bornstein & R. H. Bradley (Eds.), *Socioeconomic status, parenting, and child development* (pp. 83-106). Mahwah (NJ): Lawrence Erlbaum Associates.

Eccles, J. S., Arberton, Buchanan, Janis, Flanagan, Harold, MacIver, Midgley, Reuman. (1993). *School and family effects on the ontogeny of children's interests, self-perceptions, and activity choices.* Nebraska symposium on motivation, Lincoln, NE: U. Nebraska Press.

Elias, P., Jones, P., & McWhinnie, S. (2006). Representation of Ethnic Groups in Chemistry and Physics: The Royal Society of Chemistry and the Institute of Physics. www.raeng.org.uk/publications/other/representation-of-ethnic-groups-in-chemistry-and-p (last visited 15 October)

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.

Esminger, M. E., & Fothergill, K. (2012). A decade of measuring SES: What it tells us and where we go from here. In M. H. Bornstein & R. H. Bradley (Eds.), *Socioeconomic status, parenting, and child development* (pp. 13-28). Mahwah (NJ): Lawrence Erlbaum Associates.

Falk, J., Osborne, J., Dierking, L., Dawson, E., Wenger, M. & Wong, B., (2012). *Analysing the UK Science Education Community: The contribution of informal providers*, London, UK : Wellcome Trust.

Fang, Z. (2006). The language demands of science reading in middle school. *International Journal of Science Education, 28*(5), 491-520.

Fang, Z., & Wei, Y. (2010). Improving Middle School Students' Science Literacy Through Reading Infusion. *Journal of Educational Research, 103*, 262-273.

Finn, K. E., Yan, Z., & McInnis, K. J., (2015). Active Science: Integrating Physical Activity and Science Learning into the Afterschool Environment, *American Journal of Health Education*, 46, 323-328.

Fouche, J., (2013). The Effect of Self-Regulatory and Metacognitive Strategy Instruction on Impoverished Students' Assessment Achievement in Physics, Doctoral Thesis, Liberty University.

Furtak, E.M., Seidel, T., Iverson, H. & Briggs, D.C. (2012), Experimental and Quasi-Experimental Studies of Inquiry-Based Science Teaching: A Meta-Analysis, *Review of Educational Research*, 82(3), 300-329

Gartland, C., (2015). Student ambassadors: 'role-models', learning practices and identities, *British Journal of Sociology of Education*, 36(8), 1192-1211.

Gilleece, L., Cosgrove, J., & Sofroniou, N. (2010). Equity In Mathematics And Science Outcomes: Characteristics Associated With High And Low Attainment On PISA 2006 In Ireland. *International Journal of Science and Mathematics Education, 8*, 475-496.

Gorard, S., & See, B. H. (2009). The impact of socio-economic status on participation and attainment in science. *Studies in Science Education, 45*(1), 93-129.

Gottfried, A. W., Gottfried, A. E., Bauthurst, K., Guerin, D. W. & Parramore, M. M. (2012). Socioeconomic status in children's development and family environment: Infancy through adolescence. In M. H. Bornstein & R. H. Bradley (Eds.), *Socioeconomic status, parenting, and child development* (pp. 189-208). Mahwah (NJ): Lawrence Erlbaum Associates.

Gottfried, M. A., & Williams, D. N., (2013). STEM Club Participation and STEM Schooling Outcomes, *Education Policy Analysis Archives*, 21:79.

Griggs, M. S., Rimm-Kaufman, S. E., Meritt, E. G., & Patton, C. L., (2013). The Responsive Classroom Approach and Fifth Grade Students' Math and Science Anxiety and Self-Efficacy, *School Psychology Quarterly*, 28(4), 360-373.

Hand, B. (Ed.), (2008). *Science inquiry, argument and language: A case for the science writing heuristic*. Rotterdam, The Netherlands: Sense Publishers.

Hand, B., Norton-Meier, L. A., Gunel, M., & Akkus, R., (2016). Aligning Teaching to Learning: A 3-year Study Examining the Embedding of Language and Argumentation into Elementary Science Classrooms, *International Journal of Science and Mathematics Education*, 14, 847-863.

Hanley, P., Böhnke, J.R., Slavin, R., Elliott, L. & Croudace, T., (2016). *Let's Think Secondary Science: Evaluation report and Executive summary*, London, UK: Education Endowment Foundation.

Hanley, P., Slavin, R. & Elliott, L., (2015). *Thinking, Doing, Talking Science: Evaluation report and Executive summary*, London, UK: Education Endowment Foundation.

Hanrahan, M. (1999). Rethinking Science Literacy: Enhancing Communication and Participation in School Science through Affirmational Dialogue Journal Writing, *Journal of Research in Science Teaching*, 36(6), 699-717.

Heyneman, S. P. (1976). Influences on academic attainment: A comparison of results from Uganda and more industrialized societies. *Sociology of education*, 200-211.

Heyneman, S. P., & Loxley, W. A. (1982). Influences on academic achievement across high and low income countries: A re-analysis of IEA data. *Sociology of Education*, 13-21.

Heyneman, S. P., & Loxley, W. A. (1983). The effect of primary-school quality on academic achievement across twenty-nine high-and low-income countries. *American Journal of sociology*, 1162-1194.

Ho, E. S.C. (2010). Family Influences On Science Learning Among Hong Kong Adolescents: What We Learned From PISA. *International Journal of Science and Mathematics Education, 8*, 409-428.

Ho, H.-N. J., & Liang, J.-C. (2015). The Relationships Among Scientific Epistemic Beliefs, Conceptions of Learning Science, and Motivation of Learning Science: A study of Taiwan high school students. *International Journal of Science Education, 37*(16), 2688-2707.

Hočevar, A., Šebart, M. K., & Štefanc, D. (2013). Curriculum planning and the concept of participation in the Reggio Emilia pedagogical approach. *European Early Childhood Education Research Journal, 21*(4), 476-488.

Hoff, E. (2003). The specificity of environmental influence: Socioeconomic status affects early vocabulary development via maternal speech. *Child development, 74*(5), 1368-1378.

Hoff, E. (2012). Causes and consequences of SES-related differences in parent-to-child speech. In M. H. Bornstein & R. H. Bradley (Eds.), *Socioeconomic status, parenting, and child development* (pp. 145-161). Mahwah (NJ): Lawrence Erlbaum Associates.

Hoff, E., & Tian, C. (2005). Socioeconomic status and cultural influences on language. *Journal of Communication Disorders, 38*(4), 271-278.

Homer, M., & Ryder, J. (2015). The Impact of a Science Qualification Emphasising Scientific Literacy on Post-compulsory Science Participation: An analysis using national data. *International Journal of Science Education, 37*(9), 1364-1380. doi:10.1080/09500693.2015.1036151

Homer, M., Ryder, J., & Banner, I. (2014). Measuring determinants of post-compulsory participation in science: a comparative study using national data. *British Educational Research Journal, 40*(4), 610-636. doi:10.1002/berj.3106

Hoover, W. A., & Gough, P. B. (1990). The Simple View of Reading. *Reading and Writing: An Interdisciplinary Journal, 2*, 127-160.

Howe, C., Ilie, S., Guardia, P., Hofmann, R., Mercer, N., & Riga, F. (2015). Principled Improvement in Science: Forces and proportional relations in early secondary-school teaching, International Journal of Science Education *International Journal of Science Education, 37*(1), 162-184.

Huerta, M., Lara-Alecio, R., Tong, F. & Irby, B.J. (2014). Developing and Validating a Notebook Rubric for Fifth-Grade Non-Mainstream Students, *International Journal of Science Education*, 36(11), 1849-1870.

Inhelder, B. & Piaget, J. (1958). The growth of logical thinking from childhood to adolescence. New York: Basic Books.

Johnson, C.C. & Fargo, J.D. (2014). A Study of the Impact of Transformative Professional Development on Hispanic Student Performance on State Mandated Assessments of Science in Elementary School*, Journal of Science Teacher Education*, 25(7), 845-459.

Jones, J. D., Vanfossen, B. E., & Ensminger, M. E. (1995). Individual and organizational predictors of high school track placement. *Sociology of education*, 287-300.

Kaldon, C. R., & Zoblotsky, T. A., (2014). A Randomized Controlled Trial Validating the Impact of the LASER Model of Science Education on Student Achievement and Teacher Instruction, *Society for Research on Educational Effectiveness* (SREE) Spring 2014 Conference Presentation.

Karplus, R., Karplus, E., Formisano, M., & Paulsen, A. C. (1977). A survey of proportional reasoning and control of variables in seven countries. *Journal of Research in Science Teaching, 14*(5), 411-417.

Kaur, B., (2012). Equality and social justice in teaching and teacher education, *Teaching and Teacher Education* 28, 485-492.

Kieffer, M. J. (2012). Before and after third grade: Longitudinal evidence for the shifting role of socioeconomic status in reading growth. *Reading and Writing*, *25*(7), 1725-1746.

Kjærnsli, M., & Lie, S. (2011). Students' preference for science careers: International comparisons based on PISA 2006. *International Journal of Science Education, 33*(1), 121-144.

Klahr, D., & Nigam, M. (2004). The equivalence of learning paths in early science instruction. Psychological Science, 15, 661-667.

Klein, C. A. (1971). Differences In Science Concepts Held by Children From Three Social-Economic Levels. *School Science and Mathematics, 71*(6), 550-558.

Kohlhaas, K., Lin, H.-H., & Chu, K.-L. (2010a). Disaggregated outcomes of gender, ethnicity, and poverty on fifth grade science performance. *RMLE Online, 33*(7), 1-12.

Kohlhaas, K., Lin, H.-H., & Chu, K.-L. (2010b). Science equity in third grade. *The Elementary School Journal, 110*(3), 393-408.

Korpershoek, H., Kuyper, H., & Van der Werf, G. (2015). The relation between students' math and reading ability and their mathematics, physics, and chemistry examination grades in secondary education *International Journal of Science and Mathematics Education 13*, 1013-1037.

Lara-Alecio, R., Tong, F., Irby, B. J., Guerrero, C., Huerta, M., & Fan, Y. (2012). The Effect of an Instructional Intervention on Middle School English Learners' Science and English Reading Attainment. *Journal of Research in Science Teaching, 49*(8), 987–1011.

Lareau, A. (1987). Social class differences in family-school relationships: The importance of cultural capital. *Sociology of Education*, 73-85.

Lawson, A. E. (1983). Predicting science attainment: The role of developmental level, disembedding ability, mental capacity, prior knowledge and beliefs. *Journal of Research in Science Teaching, 20*, 117–129.

Lazonder, A.W. & Harmsen, R., (2016). Meta-Analysis of Inquiry-Based Learning: Effects of Guidance, *Review of Educational Research*, 86(3), pp.681-718.

Lee, O., Buxton, C., Lewis, S. & LeRoy, K. (2006). Science Inquiry and Student Diversity: Enhanced Abilities and Continuing Difficulties After an Instructional Intervention, *Journal of Research in Science Teaching*, 43(7), 607-636.

Lee, O., Deaktor, R., Enders, C. & Lambert, J. (2008). Impact of a Multiyear Professional Development Intervention on Science Achievement of Culturally and Linguistically Diverse Elementary Students, *Journal of Research in Science Teaching*, 45(6), 726-747.

Lemke, J.L. (1990). Talking science: Language, learning, and values. Norwood, NJ: Ablex.

Lin, E., & Shi, Q. (2014). Exploring individual and school-related factors and environmental literacy: Comparing US and Canada using PISA 2006. *International Journal of Science and Mathematics Education, 12*(1), 73-97.

Lynch, P., Benjamin, P., Chapman, T., Holmes, R., McCammon, R., Smith, A., & Symons, R. (1978). Scientific language and the high school pupil: Part 1. *Research in Science Education, 8*(1), 175-182.

Lyon, E.G. (2013) What about language while equitably assessing science?: Case studies of preservice teachers' evolving expertise, *Teaching and Teacher Education*, 32, 1-11.

Maerten-Rivera, J., Myers, N., Lee, O., & Penfield, R. (2010). Student and school predictors of high-stakes assessment in science. *Science Education, 94*(6), 937-962. doi:10.1002/sce.20408

Marks, G. (2006) Are between- and within-school differences in student performance largely due to socio-economic background? Evidence from 30 countries, *Educational Research*, *48* (1), 21-40, DOI: 10.1080/00131880500498396

Marx, R. W., Blumenfeld, P. C., Krajcik, J. S., Fishman, B., Soloway, E., Geier, R., & Tal, R. T., (2004). Inquiry-Based Science in the Middle Grades: Assessment of learning in Urban Systemic Reform, *Journal of Research in Science Teaching*, 41(10), 1063-1080.

McConney, A., & Perry, L. B. (2010). Science and mathematics attainment in Australia: The role of school socioeconomic composition in educational equity and effectiveness. *International Journal of Science and Mathematics Education, 8*(3), 429-452.

McLaughlin, C., (2014). Urban science education: examining current issues through a historical lens, *Cultural Studies of Science Education*, 9(4), 885-92.

Mercer, N., Dawes, L., Wegerif, R., & Sams, C. (2004). Reasoning as a scientist: ways of helping children to use language to learn science. *British Educational Research Journal, 30*(3), 359-377.

Mere, K., Reiska, P., & Smith, T. M. (2006). Impact of SES on Estonian students' science attainment across different cognitive domains. *Prospects, 36*(4), 497-516.

Mo, J. (2016). Program for International Student Assessment (PISA). Results from 2015. Country Note: United Kingdom. Paris: OECD.

Morgan, P. L., Farkas, G., Hillemeier, M. M., & Maczuga, S. (2016). Science Attainment Gaps Begin Very Early, Persist, and Are Largely Explained by Modifiable Factors. *Educational Researcher, 45*(1), 18-35. doi: 10.3102/0013189x16633182

Morrow, L. M., Pressley, Smith, K.J & Smith, M. (1997). The Effect of a Literature-Based Program Integrated into Literacy and Science Instruction with Children from Diverse Backgrounds,. *Reading Research Quarterly, 32*(1), 54-76.

Mosatche, H.S., Matloff-Nieves, S., Kekelis, L. & Lawner, E.K., (2013). Effective STEM programs for adolescent girls: three approaches and many lessons learned, *Afterschool Matters,* Spring 2013, 17-25.

Nagy, W. E., & Anderson, R. C. (1984). How many words are there in printed school English? *Reading research quarterly*, 304-330.

Nagy, W. E., Herman, P. A., & Anderson, R. C. (1985). Learning words from context. *Reading research quarterly*, 233-253.

Neale, M. D. (1997). *Neale Analysis of Reading Ability - Revised* (Second Revised British Edition: Standardization by C. Whetton, L. Caspall, & K. McCulloch ed.). Windsor, UK.

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

Nunes, T., Bryant, P., Barros, R. (2012). The development of word recognition and its significance for later reading skills. *Journal of Educational Psychology*, *104*, 959-973. DOI: 10.1037/a0027412.

Nunes, T., Bryant, P., Evans, D., Bell, D., Gardner, S., Gardner, A., & Carraher, J. (2007). The contribution of logical reasoning to the learning of mathematics in primary school. *British Journal of Developmental Psychology, 25*, 147-166.

O'Sullivan, P. S., Niehaus, B., Lockspeiser, T. M., & Irby, D. M. (2009). Becoming an academic doctor: perceptions of scholarly careers. *Medical Education, 43*(4), 335-341.

Oliver , M., Venville, G., & Adey, P. (2012). Effects of a Cognitive Acceleration Programme in a Low Socioeconomic High School in Regional Australia . *International Journal of Science Education, 34*(9), 1393-1410.

Olsen, R. V., & Lie, S. (2011). Profiles of students' interest in science issues around the world: Analysis of data from PISA 2006. *International Journal of Science Education, 33*(1), 97-120.

Organization for Economic Cooperation and Development (OECD) (2016*). PISA 2015 Assessment and Analytical Framework Science, Reading, Mathematics and Financial Literacy* DOI:10.1787/9789264255425-en

Organization for Economic Cooperation and Development (OECD) (2016a). PISA 2015 Results. Excellence and equity in education (Vol. I). Paris. OECD.

Organization for Economic Cooperation and Development (OECD) (2016b). PISA 2015 Results. Policies and practices for successful schools (Vol. II). Paris: OECD.

Osborne, J. & Dillon, J. (Eds.), (2010). *Good Practice in Science Teaching: What research has to say,* Maidenhead, UK: Open University Press.

Osborne, J. (2010). Arguing to Learn in Science: The Role of Collaborative, Critical Discourse. *Science, 328*(5977), 463-466.

Osborne, J. F., & Patterson, A. (2011). Scientific argument and explanation: A necessary distinction? *Science Education*, *95*, 627–638.

Osborne, J. F., Erduran, S., & Simon, S. (2004). Enhancing the quality of argumentation in school science. *Journal of Research in Science Teaching, 41*, 994–1020.

Perera, L. D. H., (2014). Parents' Attitudes Towards Science and their Children's Science Achievement, *International Journal of Science Education*, 36(18),3021-3041.

Perry, L. B., & McConney, A. (2010). Does the SES of the school matter? An examination of socioeconomic status and student attainment using PISA 2003. *Teachers College Record, 112*(4), 1137-1162.

Plass, J.L., Milne, C., Homer, B.D., Schwartz, R.N., Hayward, E.O., Jordan, T., Verkuilen, J., Ng, F., Wang, Y. & Barrientos, J., (2012). Investigating the effectiveness of computer simulations for chemistry learning, *Journal of Research in Science Teaching*, 49(3), 394-419.

Plowden report: http://www.educationengland.org.uk/documents/plowden/plowden1967-1.html

Romance, N., & Vitale, M. (2016). Implications of a Cognitive Science Model Integrating Literacy in Science on Attainment in Science and Reading: Direct Effects in Grades 3–5 with Transfer to Grades 6–7. *International Journal of Science and Mathematics Education,* 1-17.

Rose, J. (2009). *The independent review of the primary curriculum: Final report*: Department for Children, Schools and Families.

Royal Society, (2008). *Exploring the relationship between socioeconomic status and participation and attainment in science education*, London, UK: Royal Society.

Ruby, A., (2006). Improving Science Achievement at High-Poverty Urban Middle Schools, *Science Education*, 90(6), 1005-1027.

Rutter,, M., Maughan, B., Mortimore, P., Ouston, J. with Smith, A. (1979). Fifteen Thousand Hours — Secondary schools and their Effects on Schoolchildren. London: Open Books.

Samarapungavan, A, Vosniadou, S, & Brewer, W. F. (1996). Mental models of the earth, sun and moon: Indian children's cosmologies. *Cognitive Development, 11*, 491-521.

Schneider, B., Judy, J. & Mazuca, C., (2012). Boosting STEM interest in high school, *The Phi Delta Kappan*, 94(1), 62-65.

Schoon, I., Ross, A., & Martin, P. (2007). Science related careers: aspirations and outcomes in two British cohort studies. *Equal Opportunities International, 26* (2), 129-143. doi:10.1108/02610150710732203

Seah, L. H., Clarke, D., & Hart, C. (2011). Understanding students' language use about expansion through analyzing their lexicogrammatical resources. , doi:10.1002/sce.20448. *Science Education, 95.*(5), 852–876.

Seah, L. H., Clarke, D., & Hart, C. (2015). Understanding Middle School Students' Difficulties in Explaining Density Differences from a Language Perspective. *International Journal of Science Education, 37*(14), 2386-2409.

Shayer, M. (1999). Cognitive acceleration through science education II: Its effects and scope. *International Journal of Science Education, 21*, 883–902.

Shayer, M. & Adey, P. S. (1992a). Accelerating the development of formal thinking in middle and high school students, II: Postproject effects on science achievement. *Journal of Research in Science Teaching, 29* (1), 81-92.

Shayer, M. & Adey, P. S (1992b). Accelerating the development of formal thinking in middle and high school students, III: Testing the permanency of effects. *Journal of Research in Science Teaching, 29* (10), 1101-1115.

Shen, C., & Tam, H.-P. (2008). The paradoxical relationship between student attainment and self-perception: A cross-national analysis based on three waves +of TIMSS data. *Educational research and Evaluation, 14*(1), 87-100.

Siegler, R. S., & Liebert, R. M. (1974). Effects of contiguity, regularity and age on children's causal inferences. *Developmental Psychology, 10*, 574-579.

Singh, K., Granville, M., & Dika, S. (2002). Mathematics and science attainment: Effects of motivation, interest, and academic engagement. *The Journal of Educational Research, 95*(6), 323-332.

Sirin, S. R. (2005). Socioeconomic status and academic attainment: A meta-analytic review of research. *Review of Educational Research, 75*(3), 417-453.

Sjøberg, S., & Schreiner, C. (2005). *How do learners in different cultures relate to science and.* Paper presented at the Asia-Pacific Forum on Science Learning and Teaching.

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.

Smith, E. (2011). Staying in the science stream: patterns of participation in A-level science subjects in the UK. *Educational Studies, 37*(1), 59-71. doi:10.1080/03055691003729161

Smith, M. K. (1941). Measurement of the size of general English vocabulary through the elementary grades and high school. *Genetic Psychology Monographs, 24*, 311-345.

Snow, C. E. (2010). Academic Language and the Challenge of Reading for Learning About Science. *Science, 328*(5977), 450-451.

Sperling, R. A., Howard, B. C., Miller, L. A., & Murphy, C. (2002). Measures of Children's Knowledge and Regulation of Cognition. *Contemporary Educational Psychology, 27*, 51–79.

Strand, S. (2014). School effects and ethnic, gender and socio-economic gaps in educational attainment at age 11. *Oxford Review of Education, 40*(2), 223-245. doi: 10.1080/03054985.2014.891980

Strand, S. (2015). Ethnicity, deprivation and educational attainment at age 16 in England: Trends over time. London: Department for Education.

Sylva, K. (2014). The role of families and pre-school in educational disadvantage. *Oxford Review of Education, 40*(6), 680-695. doi: 10.1080/03054985.2014.979581

Taskinen, P. H., Schütte, K., & Prenzel, M. (2013). Adolescents' motivation to select an academic science-related career: the role of school factors, individual interest, and science self-concept. *Educational Research and Evaluation, 19*(8), 717-733.

Tenenbaum, H. R., Rappolt-Schlichtmann, G., & Zanger, V. V., (2004). Children's learning about water in a museum and in the classroom, *Early Childhood Research Quarterly*, 19, 40-58.

Themelis, S. (2008). Meritocracy through education and social mobility in post-war Britain: a critical examination. *British Journal of Sociology of Education, 29*(5), 427-438.

Thomas, A. S., Bonner, S. M., Everson, H. T., & Somers, J. A., (2015). Leveraging the power of peer-led learning: investigating effects on STEM performance in urban high schools, *Educational Research and Evaluation*, 21(7-8), 537-557.

Tobin, K., (2005). Building Enacted Science Curricula on the Capital of Learners, *Science Education*, 89(4), 577-594.

Tong, F., Irby, B.J., Lara-Alecio, R., Guerrero, C., Fan, Y. & Huerta, M., (2014). A Randomized Study of a Literacy-Integrated Science Intervention for Low-Socio-economic Status Middle School Students: Findings from first-year implementation, *International Journal of Science Education*, 36(12), 2083-2109.

Topcu, M. S., & Yilmaz-Tuzun, O. (2009). Elementary Students' Metacognition and Epistemological Beliefs Considering Science Attainment, Gender and Socioeconomic Status *Elementary Education Online, 8*(3), 676-693.

Toth, E.E., Klahr, D. & Chen, Z. (2000) Bridging Research and Practice: A Cognitively Based Classroom Intervention for Teaching Experimentation Skills to Elementary School Children. *Cognition & Instruction, 18(4),* 423-459.

Uerz, D., Dekkers, H., & Béguin, A. A. (2004). Mathematics and Language Skills and the Choice of Science Subjects in Secondary Education, *Educational Research and Evaluation, 10*(2), 163-182.

van der Graaf, J., Segers, E., & Verhoeven, L. (2015). Scientific reasoning abilities in kindergarten: dynamic assessment of the control of variables strategy. *Instr Sci, 43*, 381–400. DOI 10.1007/s11251-015-9344-y.

Van Laere, E., Aesaert, K., & van Braak, J. (2014). The Role of Students' Home Language in Science Attainment: A multilevel approach *International Journal of Science Education, 36*(16), 2772-2794. DOI: 10.1080/09500693.2014.936327.

van Langen, A., Rekers-Mombarg, L., & Dekkers, H. (2008). Mathematics and Science Choice following Introduction of Compulsory Study Profiles into Dutch Secondary Education. *British Educational Research Journal, 34*(6), 733-745.

Wechsler, D. (1974). *Wechsler Intelligence Scale for Children - Revised*. Windsor, England: National Foundation for Educational Research.

Weinburgh, M. (1995). Gender differences in student attitudes toward science: A meta-analysis of the literature from 1970 to 1991. *Journal of Research in Science Teaching, 32*(4), 387-398.

Wellcome Trust (2014). *Primary Science: Is it missing out? Recommendations for reviving primary science*. London: Wellcome Trust.

Wellington, J. & Ireson, G., (2008). *Science Learning, Science Teaching,* Abingdon, UK: Routledge.

White, B. Y., & Frederiksen, J. R. (1998). Inquiry, Modeling, and Metacognition: Making Science Accessible to All Students. *Cognition and Instruction, 16*(1), 3-118.

White, K. R. (1982). The relation between SES and academic attainment. *Psychological Bulletin, 91*, 461-481.

Willson, V. L. (1983). A meta-analysis of the relationship between science achievement and science attitude: kindergarten through college. *Journal of Research in Science Teaching, 20*(9), 839-850.

Xu, L., & Clarke, D. (2012). Student difficulties in learning density: A distributed cognition perspective. *Research in Science Education, 42*, 769–789.

Yang, Y. (2003). Dimensions of Socio-economic Status and their Relationship to Mathematics and Science Attainment at Individual and Collective Levels,. *Scandinavian Journal of Educational Research, 47*, 21-41. doi: 10.1080/00313830308609

Ye, L., Recker, M., Walker, A., Leary, H. & Yuan, M., (2015), Expanding approaches for understanding impact: integrating technology, curriculum and open educational resources in science education, *Education Technology Research Development*, 63, 355-380.

Yerdelen-Damar, S., & Peşman, H. (2013). Relations of Gender and Socioeconomic Status to Physics through Metacognition and Self-Efficacy *Journal of Educational Research, 106*(4), 280-289.

Yerrick, R. & Johnson, J., (2011). Negotiating White science in rural Black America: A case for navigating the landscape of teacher knowledge domains, *Cultural Studies of Science Education*, 6, 915-939.

Yin, Y., Tomita, M. K., & Shavelson, R. J. (2014). Using formal embedded formative assessments aligned with a short-term learning progression to promote conceptual change and attainment in science. *International Journal of Science Education, 36*(4), 531-552.

Young, D. J. (1990). The investigation of school effects on student attainment in science: A multilevel analysis of educational data. *Research in Science Education, 20*(1), 306-315. doi:10.1007/bf02620507

Zhang, D., & Campbell, T. (2015). An examination of the impact of teacher quality and "opportunity gap" on student science attainment in China. *International Journal of Science and Mathematics Education, 13*(3), 489-513.

Zheng, B., Warschauer, M., Hwang, J. K. & Collins, P., (2014). Laptop Use, Interactive Science Software, and Science Learning Among At-Risk Students, *Journal of Science Education and Technology,* 23, 591-603.

Zimmerman, B. J., & Kitsantas, A. (2014). Comparing students' self-discipline and self-regulation measures and their prediction of academic attainment. *Contemporary Educational Psychology, 39* 145–155.

Zohar, A., & Peled, B. (2008). The effects of explicit teaching of metastrategic knowledge on low- and high-achieving students. *Learning and Instruction*, *18*, 337-353.

Zucker, A.A. & Hug, S.T., (2008), Teaching and Learning Physics in a 1:1 Laptop School, *Journal of Science Education and Technology,* 17, 586-594.

Appendix 1.1: Measuring Socio-economic Status (SES)

In the educational literature, SES is a multidimensional construct measurable through three parent related factors: income, occupational status, and educational achievement (Bornstein, Hahn, Suwalsky, & Haynes, 2012).

Income provides families with material (e.g. desk, computer, dictionary) and cultural resources (e.g. entertainment books, access to arts, travel) and has been used in some large scale studies (e.g. in PISA 2006; Gilleece, Cosgrove, & Sofroniou, 2010). Income also influences where children go to school and what other cultural resources are available to them (libraries, museums etc.), adding up to a neighbourhood effect on children's everyday lives (Leventhal & Brookes-Gunn, 2012). It can be measured as a dichotomous variable (e.g. below vs above poverty level; Tourangeau et al., 2006; entitlement to free school meals; Sammons, Nuttall, & Cuttance, 1993) or as an ordinal scale, which is more sensitive. Although convenient, categorical measures are subject to criticisms (Hobbs, & Vignoles, 2010).

Occupational status provides parents with differential income but also different ecologies, including interests, friendships and work schedules (Esminger & Fothergill, 2012). Measures are often based only on the father's occupation, which tend to be more stable throughout their children's lives and more closely associated with their educational achievement than mother's occupations (Gottfried et al., 2012). Occupational status if often measured as an ordinal scale based on some previously defined classification.

Educational achievement provides parents with different linguistic, personal (e.g. beliefs about parenting) and cultural (e.g. acquired knowledge and tastes) resources, which influence children's everyday lives. It is measured either in ordinal scales (defined by levels in the educational system) or as categories (e.g. completed secondary education or not). Mothers' education relates to the language that they use with their children (Bornstein, Hahn, Suwalsky, & Haynes, 2012; Hoff, 2003; 2012; Hoff & Tian, 2005) and to parenting beliefs and behaviours (e.g. Eccles, 1993; McLoyd, 1998). Mothers' and fathers' education are highly correlated (e.g. Kalmijn, 1991).

Composite indices are also used, such as the Hollingshead Four-Factor Index of Social Status and the International Socioeconomic Index of Occupations (Ganzeboom, De Graaf, & Treiman, 1992; Caro, & Cortés, 2012). An analysis of the relation between SES measures and children's outcomes (Gottfried, Gottfried, Bauthurst, Guerin, & Parramore, 2012) indicated that parental education was comparable to the Hollingshead index in terms of the network of correlations with children's outcomes; it can be used effectively in the absence of other indicators. The Programme for International Student Assessment (PISA) typically uses a composite index that takes into account mothers' highest level of education as well as cultural and educational resources available in the home, as reported by the participating pupils.

Appendix 1.2: Aims of Science Education in the National Curriculum for the UK and Science Literacy Measures in PISA

The National Curriculum's (NC) aims for science are "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"

(www.National%20curriculum%20in%20England%20%20science%20programmes%20of%20study%20-%20GOV.UK.htm, p. 2).

Scientific literacy in PISA

The most recent assessment, PISA 2015, focused on "science literacy, defined as 'the ability to engage with science-related issues, and with the ideas of science, as a reflective citizen'. To succeed on the PISA science test, students had to display their mastery of three skills: explaining phenomena scientifically (based on knowledge of scientific facts and ideas), evaluating and designing scientific enquiry, and interpreting data and evidence scientifically" (OECD, 2016, p.1). These components are clearly in line with the NC aims (for a sample of items, see Baldi, Jin, Skemer, Green, & Herget, 2007). The aims of the Trends in International Mathematics and Science Study (TIMMS) science assessments are also in line with the NC (for a discussion of different conceptions of scientific literacy, see Feinstein, 2011).

Appendix 1.3: School effects on participation

Homer, Ryder and Banner (2014) found a significant school effect in explaining variation in participation in science. This is particularly interesting as it would be worth knowing what does happen in schools that tends to promote greater participation in science. However, their data set does not allow one to answer this question, beyond the fact that the school effect was not explained by the possibility of remaining in the same school for the post-16 education.

In a different study about participation in science at the end of secondary school in Ireland, Smyth and Hannan (2006) attempted to shed some light on school characteristics that affect participation in science, because they found that schools contributed significantly to the prediction of participation above and beyond students' characteristics. Among other factors, they found that the offer of science subjects before they are compulsory, time tabling, the offer of science and vocational subjects that may work well together (their example was physics and metal work), and the way science was taught (whether it was seen by students as interesting and useful) were school factors that affected participation beyond students' characteristics.

Taskinen, Schütte, and Prenzel (2013) also found that two school factors were significant predictors of interest and participation in science in Germany: the amount of additional science activities included in the school curriculum and the real-life applications in science classes.

These two sets of results, therefore, are concordant with those identified by Homer, Ryder and Banner (2014), and indicate that schools can have a positive effect on participation, through the way they organise science activities in school and the way science is taught (see also Crump, Ned, & Winkleby, 2015, for the integration of schools and universities in a science programme).

Appendix 1.4: Policies that aim at widening participation in science

Several countries, such as England and the Netherlands, responded to the view that increasing post-compulsory participation in STEM subjects is key to a successful economy in the current global and technological economy by changing the nature and amount of science with which students have to engage during compulsory education. According to Homer and Ryder (2015), the major reform of the science curriculum in 2006 for 14-16 year olds was to attempt to increase participation in post-compulsory science by engaging more students in doing science before the end of compulsory education. These innovative science qualifications, called Twenty First Century Science (21CS), aimed to improve scientific literacy through the inclusion of discussions of the nature of science and socio-scientific issues as well as to prepare students for post-compulsory science. "'Ideas about science' in the 21CS curriculum include consideration of science issues with an ethical and social dimension, for example, health issues around air quality, greener energy sources. This could provide the potential for more discussion within the science classroom compared to more traditional approaches, and more opportunities for students to voice their opinions and hear those of other students" (Homer & Ryder, 2015, p. 1367).

Using the NPD, Homer and Ryder (2015) followed a cohort of students from the age of 14 to 18 to compare those who followed the 21CS curriculum with those who did not with respect to participation in science at post-compulsory education. A simple comparison between those who followed the 21CS curriculum with those who did not showed no difference in the up-take of science in post-compulsory studies: 15.66% and 15.67%, respectively.

Homer and Ryder considered that this simple comparison required further analysis, as the types of student opting for the more traditional vs the 21CS curriculum might differ. Thus they used a multi-level model analysis that took into account students' characteristics, such as attainment in science at 14 and 16, maths and English at 14, gender and socio-economic status. In these multilevel models, the impact of the 21CS curriculum was generally negative, significant in some cases because of the large number of participants, but the effect size was consistently small, but in the case of uptake of chemistry, there was no difference between the two curricula.

Although one must consider that the 21CS curriculum was still very new, and perhaps not implemented at its best in comparison to the traditional curriculum, Homer and Ryder's analysis does not suggest that it increased participation. Homer and Ryder further note two points of interest for this review. First, that once attainment was controlled for, there was no SES effect on participation. Second, there was a consistently negative effect of English attainment at age 14 on predicting participation in science. Mathematics attainment at age 14 was a positive predictor of uptake of physics and chemistry but a negative predictor of uptake of biology.

Shortly after this first reform, a policy of an entitlement to study the separate sciences at GCSE level - biology, chemistry and physics—collectively known as Triple Award (TA) was introduced in 2008 for students who had previously attained at least a B in science. Homer, Ryder, and Banner were able to assess whether these reforms had an impact on participation in science post-16 by analysing whether there was an increase in participation amongst the first cohort of students who completed a Triple Award (TA). They used the NPD for this longitudinal analysis and followed the cohort for two years; in their analysis,

participation in and completion of courses are treated as the same, because the NPD only contains information on completed courses. They included in their definition of science participation the traditional sciences (Physics, Chemistry, and Biology) as well as Applied Science and BTEC (Business and Technology Education Council) National Award and considered Psychology, Mathematics and History as comparators in order to include in the sample students who did and who did not take science post-16. In order to control for achievement, only those students who met the requirements for the TA were included in the analysis, regardless of opting for the TA. Their analysis indicates that students who opted for TA in the 14-16 period are more likely to continue to take science at A level.

They also provide data relevant to the question of the relation between SES and participation in science. In the 14-16 sample, 14% of the students were eligible for FSM, but only 5.2% of the students who met the requirement that would allow them to take the TA were eligible for FSM. This shows that fewer students from lower SES meet the criterion for opting for TA. But most of those who meet the criterion proceed to take science post-16: 4.2% of those who study science post-16 are eligible for FSM. In other words, those students from lower SES who do achieve the B grade in science at 16 participate in science courses later. FSM students, as other TA students who are by definition high achieving, are highly represented in science courses and less often represented in vocational courses.

Students eligible for FSM are not evenly distributed across the different A level qualifications. The representation of students eligible for FSM in science at A level can be summarised in four points:

- FSM students are under-represented in Physics and in Biology (and also in Mathematics, one of the comparator subjects);

- FSM students are not under-represented in Chemistry; their participation in Chemistry is actually the same as the percentage meeting the criterion for TA (5.2%);

- FSM students are slightly over-represented in Applied Sciences (6% are eligible for FSM);

- FSM students are over-represented in BTEC National Award Applied Sciences (9%).

Homer, Ryder, and Banner (2014) conjecture that the science subjects in which the representation of students eligible for FSM is higher are those that are more likely to translate into employment soon. Elias, Jones, and McWhinnie (2006) raised a similar hypothesis regarding the higher rates of participation by some ethnic minority groups in chemistry.

The Dutch case is reviewed here rather briefly, and only as a contrasting analysis. In 1998, responding to the same pressures of widening participation and increasing participation in science by different groups, Dutch law radically changed the freedom of choice subjects in secondary school. Four profiles of study were created for the final two years in secondary school: culture and society, economics and society, science and health, and science and technology. In the two science profiles - science and health, and science and technology - chemistry, pure mathematics and physics are mandatory, but the degree of mathematics and science instruction within the profiles varies. The two science profiles provide greater opportunities for admission to university than the non-science profiles.

Van Langen, Rekers-Mombarg, and Dekkers (2008) followed a cohort of about 3,500 pupils in order to identify the predictors of choice of profiles that could be measured prior to the pupils' choice. Thus, their study, differently from that by Homer, Ryan, and Banner (2014), examined what predicts the choice of profiles given that more science and mathematics are mandatory for certain profiles. In their sample, almost 63% chose a society (non-science) profile, slightly more than 21% chose a science and health profile, and 16% chose a science and technology profile. Thus, the two science profiles were less likely to be chosen by pupils than the non-science profiles. Only the science and technology profile was initially designed to prepare pupils for a subsequent STEM (science, technology, engineering and mathematics) study but adjustments were made later to give direct access to pupils with the science and health profile to most STEM studies at university level in light of educational and employment shortages.

They found significant amounts of variance in choice of profile were explained by gender, highest parental level of education, previous mathematics achievement, language achievement in the first year of secondary school (which was *negatively* correlated with choice of science profiles), secondary school track (two track systems can lead to this choice of profiles), and school.

The negative role of language achievement in the choice of science subjects was also noted in England by Homer and Ryan (2015) and was found in a previous study in the Netherlands (Uerz, Dekkers, & Beguin, 2003), but the language effect disappeared in the Van Langen, Rekers-Mombarg, and Dekkers (2008) when the measure of achievement used was obtained in later in secondary school (in the third year rather than the first year); at this time, it had no significant correlation with subsequent choice of science subjects.

The final predictive model showed that previous achievement in mathematics is the strongest predictor of choice of a profile that includes more mathematics and science; parental level of education and gender are independent predictors of choice of science profiles, but there is also an interaction between these two background variables: boys and children from more educated parents choose a science profile more often than girls and children from less educated parents. Similarly to the findings in the Homer, Ryan and Banner (2014) study, schools contributed significantly to the choice of a science profile.

In brief, the analysis of the outcomes of the policy of increasing the amount of science required during compulsory education does not suggest that this policy can counteract the SES effects. The UK policy assessment shows that, once achievement is controlled for, the effect of SES on participation seems to be small and manifests itself more in the choice of science than participation, whereas the Dutch policy analysis shows that the effect of SES continues to be significant. However, it is worth noticing that, in the study of the UK policy, SES is defined by eligibility for FSM, a categorical variable, whereas in the Dutch study SES is defined by highest level of parental education, an ordinal scale. The difference between these measures of SES might account for the differences in results, as a more sensitive measure is more likely to find a significant SES effect.

Appendix 1.5: Intervening variables: mediators and moderators

Causal relations involve more than just an association between two variables: they require an explanation for how the cause leads to the effect. In the case of SES and science achievement, it is most likely that other variables are involved in the link SES and science achievement. These intervening variables are called either *mediators* or *moderators*, depending on the part that they play in the causal chain. Baron and Kenny (1986) define mediators as entities or processes that intervene between the cause and the effect. Mediators are distinct from moderators; moderators are entities or processes that affect the magnitude or the nature of the association between the measure of the cause and of the effect. Plausible mediators of the link between SES and science achievement are the parents' use of more academic vocabulary in their interaction with their children or parental practices in reading with their children, which could affect children's reading comprehension of scientific texts. Plausible moderators of the link between SES and science and educational policies that expose children to more scientific activities in school.

Translating these conceptual distinctions into predictions in statistical models, Baron and Kenny assert that "in general, a given variable may be said to function as a *mediator* to the extent that it accounts for the relation between the predictor and the criterion. Mediators explain how external physical events take on internal psychological significance. Whereas moderator variables specify when certain effects will hold, mediators speak to how or why such effects occur" (Baron & Kenny, 1986, p. 1176). Thus testing whether X mediates the relation between SES and science achievement requires examining whether the association weakens significantly or disappears when X is included in a hierarchical regression model, whereas testing whether X moderates the relation between SES and science achievement requires examining whether there is an interaction between SES and X.

Schools can be both mediators and moderators of the link between SES and science attainment. If a school's SES (which is the average of the SES of students in the school) is related, for example, to the standards of teacher qualification or to levels of resource for science teaching, the school's SES is a mediators of the link between SES and science attainment. Schools can also moderate the SES and science achievement relation: schools with high standards can modify the magnitude of the association between SES and science attainment by improving attainment for all pupils. Both hypotheses can be tested using regression analysis.

Appendix 2.1: FSM and EVER6 by Year Group January 2013

Overall in 2013 18.2% of pupils aged 5-16 were entitled to FSM and 27.6% were EVER6. The two proportions are similar for Reception year but then diverge as the cumulative nature of EVER6 comes into play through to about Y6, beyond which the EVER6 proportion is fairly stable, as illustrated in Table 1.

Table 1 Percentage of pupils entitled to FSM and EVER6 by Year Group 2013

Year group	Number students in year group	% Entitled in Jan. Census (FSM)	% entitled in last 6 years (EVER6)	Difference
Reception	635,812	19.0%	19.9%	1.0%
Y1	615,919	19.7%	24.0%	4.3%
Y2	596,582	20.0%	26.8%	6.8%
Y3	583,202	19.7%	28.6%	8.9%
Y4	571,297	19.0%	29.4%	10.4%
Y5	554,717	18.8%	30.3%	11.5%
Y6	535,033	18.5%	30.8%	12.3%
Y7	534,959	18.3%	30.8%	12.5%
Y8	545,967	17.7%	29.9%	12.2%
Y9	560,687	16.7%	28.7%	12.0%
Y10	567,412	15.9%	27.7%	11.8%
Y11	571,906	14.9%	26.6%	11.7%
All pupils	6,873,493	18.2%	27.6%	9.4%

Notes: Author's analysis of spring 2013 School Census.

Appendix 2.2: An explanation of effect size measures

We are interested in this report in establishing the size of the differences in outcomes between students who have been entitled to FSM at some time in the last 6 years (EVER6) and those who have not been entitled to a FSM at any time during that period (NonFSM).

Where the outcome of interest is in a readily interpretable or meaningful scale this can be relatively straightforward. For example, if the outcome were average income we might feel this metric is of itself meaningful. For example if the average weekly earnings of EVER6 students was £160 and the average weekly earnings of NonFSM students was £200 then NonFSM students on average earn £40 per week more than EVER6 students. However metrics in educational research are often not inherently meaningful in this way.

Cohen's D

Suppose rather than £, Kg or cm we are measuring achievement as indicated by GCSE average points score. It can be difficult to interpret what constitute a large or a small gap in terms of points scores. It is also difficult to compare the size of the gap in GCSE points score at age 16 with the size of the gap measured in National Curriculum levels at age 7, or in KS2 test marks at age 11.

One way to estimate the absolute size of the gap, and to do this in a form that is consistent across many different measures, is to calculate Cohen's D. Cohen's D is an effect size measure for use with continuous variables. It is calculated as:

$$\textit{Cohen's D } = \frac{[\text{Mean of comparator group}] - [\text{Mean of reference group}]}{\text{Pooled Standard Deviation}}$$

There is no restriction on which is the comparator and which the reference group as the absolute value of the difference between the two groups is the same whichever is defined as the reference group, though the sign of the difference (+/-) will change. The important thing is this expresses the difference between the groups in standard deviation (SD) units. The 'standard deviation' is a measure of the spread of a set of values and here it refers to the pooled standard deviation of the whole sample. The interpretation is therefore consistent whatever units the outcome is measured in since the Cohen's D gives the gap as the number of SD units, and so is comparable across many different measures.

Cohen's D effect sizes are generally given labels of "small", "moderate", or "large". The most frequent guidelines from Cohen (1988) are 0.2 is small, 0.5 is medium and 0.8 is large. However these are rough guidelines not cut-off values. However what constitutes a small, moderate, or large effect does depend on the area of research and should be interpreted relative to typical results in the particular field of enquiry.

Odds Ratios

Odds Ratios are an effect size measure used where the outcome is dichotomous or binary, for example a student achieves 5 or more GCSE passes at A*-C including English and Mathematics (5EM) or does not achieve this outcome. We can report these percentages achieving the outcome for the two groups and the figures can be directly compared. For example we saw in Table X of the report that 33.1% of pupils entitled to FSM achieve the

5EM threshold compared to 60.9% of pupils not entitled to FSM. However we sometimes want to go further, we might want to:

(i) compare across different measures, for example if 10% of EVER6 and 20% of NonFSM achieve outcome X, how does this gap compare in size to the gap for outcome Y which is achieved by 50% of EVER6 and 60% of NonFSM students? (ii) compare changes in the percentages achieving a particular measure over time, for example if the proportion of EVER6 students achieving outcome X increases from 10% to 30% and the proportion of NonFSM increases from 25% to 50% has the gap widened, closed or stayed the same? (iii) Explore how other variables may impact on or change the probabilities of the outcome occurring for the two groups through a technique called logistic regression.

For these reasons the Odds Ratio (OR) is a particularly useful effect size measure. The OR compares the odds of the outcome occurring for the comparison group (say EVER6) divided by the odds of the outcome occurring for a reference group (say NonFSM). The OR can range from 0 to infinity where:

- OR >1 indicates the odds of the outcome occurring are higher for the comparison group relative to the reference group
- OR =1 indicates the odds of the outcome occurring are equal for both groups
- OR <1 indicates the odds of the outcome occurring are lower for the comparison group relative to the reference group

The OR is contingent on which group is defined as the reference group. For example if the odds of an outcome are twice as high for girls as boys (OR=2.0) this is equivalent to saying the odds of the outcome are half as high for boys as for girls (OR=0.50). The ratios are equivalent, they just vary depending on whether it is the boys or the girls who are defined as the reference group. Any OR can be converted to its complement by dividing the OR into 1 (e.g. 1/2 = 0.50, and 1/0.5= 2.0).

To illustrate the process, consider the odds of achieving Level 2 or above for KS1 reading (see Chapter 2, p26). The odds for NonFSM pupils achieving this threshold are .931/(1-.931)=13.5. The odds for EVER6 pupils achieving this threshold are .838/(1-.838)=5.2. So the ratio of the two odds (the odds ratio) is 2.6.

We should note that effect sizes do not imply causality, they are just a measure of the association between two variables.

For further references on Effect sizes see Cohen (1988) and Coe (2004).

Appendix 2.3: Achievement and progress during Key Stage 4 by subject and EVER6: 2014 and 2013

2014	Never Entitled FSM			Entitled FSM last 6 years			All pupils			Odds Ratio
KS4 Indicator	Mean	N	SD	Mean	N	SD	Mean	N	SD	(Cohen's D)
Attempted Science GCSE	**86.6%**	412,458	0.34	**70.1%**	152,848	0.46	**82.1%**	565,306	0.38	2.8
Achieved A*-G in Science GCSE	**86.2%**	412,458	0.34	**68.3%**	152,848	0.47	**81.4%**	565,306	0.39	2.9
Achieved A*-C in Science GCSE	**68.1%**	412,458	0.47	**39.5%**	152,848	0.49	**60.4%**	565,306	0.49	3.3
Achieved equivalent of Level 2 in BTECs in Science	**12.1%**	412,458	0.33	**22.1%**	152,848	0.42	**14.8%**	565,306	0.36	0.48
Highest point score in Science (GCSE & equiv.)	**42.8**	361,652	9.32	**36.7**	109,288	10.87	**41.4**	470,940	10.04	0.61
Entered for 3 individual sciences	**26.2%**	412,458	0.44	**10.3%**	152,848	0.30	**21.9%**	565,306	0.41	3.1
Entered EBacc (minimum two sciences)	**74.9%**	412,458	0.43	**49.1%**	152,848	0.50	**67.9%**	565,306	0.47	3.1
Achieved two Sciences A*-G	**74.7%**	412,458	0.43	**48.5%**	152,848	0.50	**67.6%**	565,306	0.47	3.1
Achieved EBacc (core & additional or double)	**32.4%**	412,458	0.47	**19.9%**	152,848	0.40	**29.0%**	565,306	0.45	1.9
Achieved EBacc (three separate pathway)	**24.5%**	412,458	0.43	**8.8%**	152,848	0.28	**20.3%**	565,306	0.40	3.4
Achieved EBacc (total)	**56.6%**	412,458	0.50	**28.5%**	152,848	0.45	**49.0%**	565,306	0.50	3.3
EBacc Science points score (0 if none)	**32.5**	412,458	20.07	**19.2**	152,848	20.51	**28.9**	565,306	21.03	0.63
Best 8 points score (GCSE & equiv.)	**327.2**	412,458	90.2	**253.4**	152,848	113.3	**307.3**	565,306	102.4	0.72
Included in Science VA calculation	**72.1%**	412,458	0.45	**47.4%**	152,848	0.50	**65.4%**	565,306	0.48	2.9
EBacc Science VA score	**0.34**	297,417	5.85	**-1.40**	72,427	6.91	**0.00**	369,844	6.11	0.28
Best 8 VA Score (all pupils)	**23.0**	412,458	104.3	**-11.1**	152,848	115.5	**13.7**	565,306	108.5	0.31
Best 8 VA Score (for EBacc Science pupils)	**17.4**	297,430	51.1	**6.5**	72,446	64.4	**15.3**	369,876	54.15	0.20

Notes: *See footnotes to Table 2.3.*

2013	Never Entitled FSM			Entitled FSM last 6 years			All pupils			Odds Ratio
KS4 Indicator	Mean	N	SD	Mean	N	SD	Mean	N	SD	(Cohen's D)
Attempted Science GCSE	**84.1%**	419,052	0.37	**65.1%**	152,084	0.48	**79.1%**	571,136	0.41	2.8
Achieved A*-G in Science GCSE	**83.8%**	419,052	0.37	**63.7%**	152,084	0.48	**78.4%**	571,136	0.41	2.9
Achieved A*-C in Science GCSE	**65.5%**	419,052	0.48	**36.8%**	152,084	0.48	**57.9%**	571,136	0.49	3.3
Achieved equivalent of Level 2 in BTECs in Science	**0.1%**	419,052	0.03	**0.1%**	152,084	0.02	**0.1%**	571,136	0.03	n.a.
Highest point score in Science (GCSE & equiv.)	**43.1**	352,768	9.1	**37.7**	97,263	9.7	**41.9**	450,031	9.5	0.57
Entered for 3 individual sciences	**29.3%**	419,052	0.45	**12.3%**	152,084	0.33	**24.7%**	571,136	0.43	3.0
Entered EBacc (minimum two sciences)	**73.2%**	419,052	0.44	**47.2%**	152,084	0.50	**66.3%**	571,136	0.47	3.1
Achieved two Sciences A*-G	**73.0%**	419,052	0.44	**46.5%**	152,084	0.50	**65.9%**	571,136	0.47	3.1
Achieved EBacc (core & additional or double)	**28.5%**	419,052	0.45	**17.3%**	152,084	0.38	**25.5%**	571,136	0.44	1.9
Achieved EBacc (three separate pathway)	**26.8%**	419,052	0.44	**10.1%**	152,084	0.30	**22.4%**	571,136	0.42	3.3
Achieved EBacc (total)	**55.5%**	419,052	0.50	**27.6%**	152,084	0.45	**48.1%**	571,136	0.50	3.3
EBacc Science points score (0 if none)	**31.8**	419,052	20.5	**18.48**	152,084	20.49	**28.3**	571,136	21.3	0.63
Best 8 points score (GCSE & equiv.)	**354.7**	419,052	72.3	**306.2**	152,084	96.8	**341.8**	571,136	82.4	0.59
Included in Science VA calculation	**70.7%**	419,052	0.46	**45.6%**	152,084	0.50	**64.0%**	571,136	0.48	2.9
EBacc Science VA score	**0.56**	296,192	6.01	**-1.30**	69,329	7.20	**0.21**	365,521	6.29	0.29
Best 8 VA Score (all pupils)	**4.1**	419,052	58.0	**-19.7**	152,084	88.4	**-2.3**	571,136	68.3	0.35
Best 8 VA Score (for EBacc Science pupils)	**9.4**	296,234	45.6	**-0.6**	69,349	56.9	**7.5**	365,583	48.1	0.21

Notes: See footnotes to Table 2.3.

Appendix 2.4: KS4 (age 16) EBacc science outcomes by pupil and school characteristics

		EBacc two Sciences points score (0 if none)			Entered EBacc Science
		Mean	**N**	**SD**	
KS2 prior attainment band	Below Level 4	9.3	92,963	14.8	30.0%
	Level 4	29.5	257,466	17.9	75.4%
	Level 5 or above	45.0	82,044	12.0	95.3%
	No KS2 score	24.3	28,379	22.2	57.7%
Gender	Girls	32.1	273,579	19.7	75.6%
	Boys	29.7	287,273	20.2	71.4%
Ethnic Group	White Irish	35.4	1,903	19.2	80.0%
	Traveller Irish	10.7	163	17.0	30.7%
	Traveller Gypsy/Roma	7.5	1,097	14.3	25.6%
	White other groups	28.1	22,866	21.3	66.6%
	Mixed White & African	31.0	2,569	19.9	73.9%
	Mixed White & Caribbean	27.2	7,519	20.1	67.6%
	Mixed White & Asian	35.6	4,911	19.3	80.4%
	Mixed other background	32.5	8,010	20.2	75.3%
	Indian	38.1	13,384	18.4	83.7%
	Pakistani	29.0	20,099	20.5	69.5%
	Bangladeshi	32.6	8,075	19.7	76.0%
	Any other Asian	35.5	8,230	19.7	79.3%
	Black African	30.7	16,879	19.9	73.2%
	Black Caribbean	26.7	7,529	19.6	67.7%
	Black other groups	26.1	3,314	20.5	64.6%
	Chinese	41.5	2,078	17.9	86.6%
	Any other ethnic group	30.8	7,900	20.9	71.4%
	Unclassified/Refused	29.1	5,865	20.8	69.8%
	White British	30.9	418,461	19.8	73.8%
EAL	First language English	31.1	481,090	19.8	74.0%
	English Additional Language	29.9	78,764	20.8	70.4%
FSM	No FSM Jan 2015	32.7	481,696	19.4	76.8%
	Eligible FSM Jan 2015	20.0	79,156	20.1	52.8%
IDACI deprivation Quintile	1 Very low	37.9	111,950	17.5	85.1%
	2 Low	34.6	111,938	18.6	80.5%
	3 Average	30.9	111,935	19.6	74.3%
	4 High	27.0	111,937	20.3	66.7%
	5 Very high	24.2	111,903	20.7	60.8%
SEN stage	No identified SEN	34.4	465,058	18.4	80.7%
	SEN (Unstatemented)	16.0	73,789	19.2	44.1%
	SEN (Statemented)	6.7	22,005	14.6	18.7%
Mobile	Same school throughout	31.3	543,130	19.8	74.3%
	Joined in Y10-Y11	17.1	17,722	19.6	46.4%

		EBacc 2 Sciences points score (0 if none)			Entered EBacc Science
		Mean	**N**	**SD**	
School type	LA maintained	31.4	108,683	19.2	75.7%
	Voluntary aided/controlled	33.5	54,207	19.0	78.3%
	Foundation/CTC	29.2	53,570	19.8	71.5%
	Academy-Sponsored	24.6	79,448	20.3	62.6%
	Academy-Converter	34.7	244,012	18.8	80.4%
	Special	0.5	10,724	4.1	1.4%
	Free school (UTC, Studio)	28.5	3,247	18.3	74.5%
	PRU/Secure Unit/AP	0.6	1,188	4.8	2.1%
	FE College	0.9	5,631	5.7	2.8%
School Selective Status	Comprehensive	31.3	501,040	19.4	75.2%
	Selective	47.8	22,492	14.2	93.2%
	Secondary Modern	27.8	19,302	19.5	69.6%
	Other	1.2	18,018	6.6	3.3%
Region	North East	28.5	27,333	20.5	68.4%
	North West	29.6	77,983	20.5	70.4%
	Yorkshire & Humberside	29.4	57,436	20.2	71.2%
	East Midlands	29.9	49,631	20.0	72.1%
	West Midlands	30.4	62,373	19.9	73.0%
	East England	31.6	61,112	19.4	75.8%
	Inner London	30.8	21,703	20.4	72.1%
	Outer London	32.7	56,521	20.1	75.5%
	South East	32.2	92,108	19.7	75.9%
	South West	32.0	54,652	19.4	76.0%

Notes
EBacc two sciences= Mean GCSE points score of the best two eligible EBacc sciences subjects, with 0 entered for a missing score if there is one and 0 for those not entered for any EBacc science subjects at all. EAL= English as an Additional Language; LA= Local Authority; CTC= City Technology College; UTC= University Technical College, PRU= Pupil Referral Unit; AP=Alternative Provision; FE= Further Education.

Appendix 2.5 - Methodology for the analysis of participation and achievement at the end of KS5 (Age 19)

Eligible students in 2015

Students were eligible to be reported in the 16-18 performance tables in 2015 if they satisfied the following criteria:

1. Were aged 16, 17 or 18 on 31 August 2014
2. Were recorded on the school roll in January 2015
3. Were in, or deemed to be in, Year 13
4. Completed their advanced studies in the 2014/15 academic year
5. Entered for an A level or substantial level 3 qualification equal in size to an A level in the reporting year (180 guided learning hours per year).

In 2015 404,000 pupils were eligible for inclusion in the 16-18 tables[18].

Missing FSM data

Information of whether or not a student was entitled to a FSM was only available for 182,467 students, or just 45% of the above sample. The 54.9% with missing FSM data were not just the 37,246 attending Independent schools, which is to be expected, but 65,706 students attending Sixth form Colleges and 117,901 attending FE Colleges. The absence of any FSM data from such a large proportion of students in state-funded post-16 establishments renders a direct analysis of this data problematic.

Addressing the problem - matching to the KS4 cohort

To address the missing FSM data problem, we decide to use the EVER6 indicator from when the students were in Y11 where we have complete census data. However a match of the 2015 KS5 results to the KS4 data two years earlier in 2013 revealed a further problem, as 75,675 (18.7%) of students were not matched. The vast majority of these students turned out to be older, that is though they were in 'actual' Y13 in 2015 they would be predicted to be in Y14 by their age. A further match against the KS4 dataset from a year earlier (2012) located the vast majority (65,675) of these students. Put another way, around 17% of an age cohort take three years rather than two years to complete their Level 3 programmes of study.

Given this finding, the appropriate analytic strategy is not to focus on the KS5 results in a particular year such as 2015. The more comprehensive approach is to focus on the cumulative achievement of a particular cohort of students over time.

Establishing the Age 19 cohort

In relation to the current analysis this means that using the most recently available data we:

1. Base the analysis on the KS4 cohort in summer 2012
2. Match in the KS5 results from 2014 (two years later at age 18)

---

[18]. As part of the performance tables checking exercise, institutions are able to request that their students' results are deferred from the performance tables by one year if they have not reached the end of 16-18 study. The following year, these students will be added back into institution figures to count towards that year's results, regardless of whether they have achieved any further results. A student cannot be deferred at academic age 18 as they will be outside the scope of performance tables the following year.

3. Match in the KS5 results from 2015 (three years later at age 19)

A similar process is followed by the DFE when calculating the age 19 L2/L3 data (e.g. DFE SFR 11/2015). Table 1 below summarises the matching process. Overall 61% of the KS4 2012 cohort entered for one or more level 3 qualifications by age 19.

Table 1 Matching process to establish the age 19 cohort in 2015

	All students at KS4		Maintained Mainstream schools at KS4	
	Number of students	% of KS4 cohort	Number of students	% of KS4 cohort
Finished Y11 in 2012	631,330	100.0%	551,333	100.0%
Entered KS5 Level 3 qualification in 2014	323,914	51.3%	285,356	51.8%
Entered KS5 Level 3 qualification in 2015(a)	61,306	9.7%	57,347	10.4%
Total matched KS5 sample	385,220	61.0%	342,703	62.2%
EVER6 in Y11 (i.e. Elegible FSM Y6-Y11)	-	-	134,795	24.4%

Note: (a) These are unique cases who did not complete any level 3 qualifications until 2015, i.e. not previously identified as a result of the 2014 match. A small number of students (2,617) had also been matched in 2014 so presumably were doing retakes in 2015. Because the DFE discounting rule applied to the NPD dataset only improved grades would be counted.

EVER6 data was not available for all KS4 students. EVER6 data was missing for the 47,125 pupils attending Independent schools, the 7,107 pupils in Pupil Referral Units and the 6,004 pupils in FE Colleges at KS4 (these three establishment types accounting for 91% of all missing FSM cases). We decided therefore to filter all our results to Maintained Mainstream schools to provide a consistent base for interpretation. This matching data for this group are presented in the two right-hand columns of Table 4.1.

Of the 551,333 students in Maintained Mainstream schools at KS4 for whom FSM records were available, 134,795 (24.4%) were EVER6 in Y11, i.e. had been entitled to a FSM sometime during the period Y6-Y11. Subsequent analyses compare the achievement of this EVER6 group against those who had not been entitled to FSM at any time during secondary school.

Base for Analysis
An accurate assessment of the EVER6 gap at age 19 can now be established and we can evaluate the EVER6 gap at age 19 from three bases:

- KS4 cohort: The number of students in the cohort at the end of Y11. This base gives the direct proportion of EVER6 vs. NonFSM pupil at age 16 who go on to achieve Level 3 qualifications by age 19.

- Level 3 cohort: The number of students entered for Level 3 'A' level or vocational qualifications by age 19. This base gives the EVER6 vs. NonFSM gap conditional on having studied and entered a level 3 qualification.

- 'A' level cohort: The number of students entered for at least one 'A' level or applied 'A' level qualification by age 19. This base gives the EVER6 vs. NonFSM gap conditional on having studied and entered for at least one 'A' Level.

Appendix 3.1: Details of the method used in the literature review reported in Chapter 3

A literature search was first carried out with the aim of identifying research that includes data both on SES and on science learning, as well as an explicit hypothesis about what might cause the association between SES and science learning. The search for papers relating to SES effects on science outcomes used the British Education Index, PsycINFO, which is maintained by the American Psychological Association, and ERIC (Educational Resources Information Center), a digital library sponsored by the Institute of Education Sciences of the US Department of Education.

PsycINFO provides citations with abstracts to the scholarly literature in the psychological, social, behavioural, and health sciences. It contains more than 2 million records spanning 1806 to the present. Journal coverage includes material selected from approximately 2,000 periodicals. Chapter and book coverage includes worldwide English-language material published from 1987 to the present, but there is also a substantial number of records covering books published earlier. Dissertations constitute approximately 12% of the database. The American Psychological Association, which maintains it, is currently adding approximately 8,100,000 references annually through weekly updates.

Information in the ERIC database corresponds to two printed abstract index journals: Resources in Education (RIE) and Current Index to Journals in Education (CIJE). These include records of many different types: journal articles, books, theses, curriculi, conference papers, standards and guidelines, research/technical reports, conference papers, program descriptions, opinion papers, bibliographies, state-of-the-art reviews, legal/legislative/regulatory materials, dissertations, classification schemes, teaching guides, curriculum materials, lesson plans, course descriptions, pamphlets, guides, and many other types of material. The records span 1966 to present; it is updated monthly.

Although there is overlap between these different sources, the overlap is smaller than one might expect. This search was complemented by searches through four science education journals (International Journal of Science Education, Science Education, Research in Science Education, and International Journal of Science and Mathematics Education) using the relevant terms. A total of 823 abstracts were read; 86 were read by two reviewers, who classified them as relevant, potentially relevant, or irrelevant; 5 papers were classified as potentially relevant by one reader and irrelevant by the other; this indicates a very high level of agreement. Then 142 abstracts were selected for consultation of the full papers. When the full papers were read, further papers were discarded because they did not meet the criteria; 81 papers were included in the review at this stage, but not all of them are cited for various reasons (e.g. low quality data, data sets from largely different contexts, such as African or Latin American countries). These papers were analysed to obtain the overview of the effects of SES on science outcomes, described in Chapter 3.

Appendix 4.1: Searching the literature for studies of the mediators of the relationship between SES and science learning

In order to write about possible mediators of the relation between SES and science attainment in school we did a new search in which we entered terms for possible explanations for science attainment (scientific reasoning, metacognition, conceptual change, argumentation, motivation) and SES. This produced 667 results. Once the abstracts were read, papers were selected for inclusion in the review if they included data on correlational or intervention studies that link the possible mediator to science learning.

The search for a causal mediator also had to operate in a different way: we looked for papers that identified cognitive developmental and educational outcomes that are related to SES (Bornstein, Hahn, Suwalsky & Haynes, 2012; Bradley & Corwyn, 2002; Duncan & Magnusson, 2012; Gottfried, Gottfried, Bauthurst, Guerin & Parramore, 2012; Sirin, 2005; White, 1982) and then searched the literature for evidence that these factors could have an impact on science outcomes. For example, much observational and correlational research has shown that the quantity and complexity of the vocabulary used by mothers with their children varies with the mother's level of education (see, for example, Hoff, 2003; 2012; 2013; Hoff & Tian, 2005). It is also known that SES is related to children's literacy and mathematics learning as well as to the development of the executive function (e.g. Aram, & Levin, 2001; Ardila, Rosselli, Matute, & Guajardo, 2005; Dilworth-Bart, 2012; Sirin, 2005; White, 1982) .

By themselves, these results do not signify that language, literacy, mathematics and executive function differences play a causal role in the mediation of the connection between SES and science attainment. In order for these differences to be plausible mediators of the relation between SES and science attainment, one would need to identify research which shows that language and literacy skills play a role in learning science. Thus a third search was carried out using the terms "vocabulary", "literacy", "mathematics" and "executive function" together with "science attainment" to assess whether these SES-related cognitive and educational outcomes might be mediators of the relation between SES and science attainment. This search produced 1695 results;[19] only one of which was related to executive function, and so we decided to exclude executive function from further analysis. We examined the remaining 1694 abstracts and identified papers to include in the review by cross-checking the outcomes of this new search with those of the previous one. For example, although mathematics attainment is associated with SES and also correlates with science attainment, we were unable to find papers that connected mathematics attainment with science outcomes in intervention studies, and so mathematics attainment was not included in the list of possible mediators of the SES impact on science attainment. This process of matching possible mediators identified across the searches resulted in the inclusion of just over 460 papers in the next step of the review. Not all these papers are cited in this report because we then excluded papers that were themselves reviews, or did not include data or reported on interventions on factors that were not related to SES (e.g. Yin, Tomita, & Shavelson, 2014, report a good quality study on the effect of the use of formative assessment on conceptual change, but a teacher's decision to use formative

---

[19] PsychINFO was excluded from this search because of the number of references it produced (54,030). The search is restricted to the British Educational Index.

assessment cannot be a mediator of the relation between pupils' SES and science attainment). Chapter 4 analyses three possible mediators of the impact of SES on science attainment: scientific reasoning, literacy and metacognition.

It seemed to us that another plausible candidate for consideration as a mediator of SES differences in science learning was the extent of pupils' language skills, such as their vocabulary and their ability to learn and understand new words. However, we found that most of the research relating language to science learning had concentrated, quite rightly, on written rather than oral language, and that there was very little systematic correlational work on the connection between oral language and science learning. There were several intervention studies purportedly about the effects of improving children's use of oral as well as written language in science classes, such as the highly successful intervention project by Mercer, Dawes, Wegerif, and Sams (2004), but quite naturally these interventions in school science classes involved a great deal more than just increasing the pupils' language skills. Normally the impact of the different factors in these necessarily complex interventions could be sorted out in accompanying longitudinal correlational research, but this kind of research does not seem to have been done on the possible relation between pupils' language skills and science learning. We are not making a general critical point about intervention studies here: our aim is to advocate the combination of intervention with longitudinal correlational studies.

Appendix 5.1: Hierarchical Regressions and Their Use in the Test of Mediators

When two variables are correlated, the amount of variance that they have in common is the square of the correlation. If two variables, such as individual SES and school SES are correlated, it is possible to calculate the multiple correlation between the two variables together and the outcome, which in our analyses is always KS2 or KS3 science attainment. The multiple correlation of individual and school SES with KS science attainment is not the sum of the individual correlations, because it takes into account that there is overlap between individual and school SES.

Hierarchical regression analyses allow a researcher to calculate one correlation first (the correlation between individual SES and KS science attainment) and the amount of variance it explains, and then calculate how much more variance the second measure explains (in this case, the correlation between school SES and KS science attainment). When a measure is entered in the *second* step in the regression, the amount of variance that it explains in the variable of interest (here, science attainment) is independent of the variance that this second measure and the first one have in common. If individual SES were to be the only aspect of SES that affects science attainment, entering the measure of school SES second would not explain further variance. If school SES does explain further variance in science attainment, this means that schools with a higher average SES create better environments for learning science, independently of the individual pupils' SES. This would set the stage for further research: what did schools with higher SES do that promoted their science attainment?

When the analyses are run in both sequences, i.e. with individual SES first and school SES second and then with school SES first and individual SES second, it is possible to test whether each of these measures makes an independent contribution to explaining variance in science attainment. The expression "independent contribution to explaining variance" is used in this report with this specific meaning.

Appendix 6.1: Methodology for the systematic review of promising educational approaches

We employed multiple methods to search the literature to ensure a thorough coverage and to minimise the possibilities of omitting any promising research from the review. Initially, the British Education Index (EBSCO) was searched using the following search terms:

"(low income OR low socioeconomic OR socio* OR deprivation OR poverty OR poor OR economically disadvantaged OR urban OR rural OR Xhosa OR free lunch OR reduced lunch OR free meal OR reduced meal)

AND

Science Education OR Biology OR Chemistry OR Physics

AND

1. intervention
2. program*
3. "out of school" time OR OST OR "After school" OR Clubs OR Activities OR Clinics OR Workshops pedagog* OR strategy OR teaching"

One example of a Boolean phrase would be:

(low income OR low socioeconomic or socio* OR deprivation OR poverty OR poor OR economically disadvantaged OR urban OR rural OR Xhosa OR free lunch OR reduced lunch OR free meal OR reduced meal) AND science education AND (pedagog* OR strategy OR teaching)

This process was repeated for Education Resources Information Centre (ERIC) and Education Abstracts (EA), and was then followed by searching individual leading international journals in science education and in teacher education, again using these search terms. The searches yielded nearly 5000 articles whose bibliographic details and abstracts were inputted into the EPPI reviewer software, produced by the Evidence for Policy and Practice Information Centre (EPPI-Centre) at the Institute of Education, University College London. This software offers an online tool for systematic management and rigorous analysis of reference yielded by this type of comprehensive literature search.

 From this point, each entry was included or excluded according to a number of criteria:

- The age of the study – for the disadvantaged pupils' section of the review, articles dating back to 1990 were included. Although more recent research would seem to be more relevant to the current situation in science education, and over 70% of the literature discussed has been published since a previous review by the Royal Society (2008), we were anxious not to exclude anything promising simply because it was slightly older. Extending back beyond 1990 did not seem profitable, given the significant changes to the majority of educational systems around the world since the 1980s. However, when we considered literature across the whole of science education, we divided the research was into two parts: UK research published in the last 10 years (2006 onwards) and international research published since 2011. These dates were chosen for two reasons: one is pragmatic in terms of keeping things manageable, the other is that these dates capture the time periods of significant change (to curricula and assessments) in the UK, USA, Sweden, Australia, Singapore

and South Africa (not an exhaustive list). This, however, proved to be extremely challenging, and resulted in a search focused on meta-analyses and literature reviews.

- Nature of publication – studies were included if they were published in a double-blind, peer review journal, although some national level reports and doctoral theses were also included if they met all the other criteria.
- Type of study – preference was given to studies with a strong research design and/or robust statistical analyses of findings. When considering the evidence presented by the literature, typically, more weight was given to studies with a pre- and post-test design and a control group over a study with a small number of purely qualitative case studies. However, the insights afforded by qualitative research should not be underestimated, and mixed methods studies were often found to be of considerable value.
- Target group/participants – a significant challenge was presented by this aspect of the searching, as 'disadvantaged' or 'low SES' pupil is not an well-defined parameter, hence the long list of plausible alternative search terms in the initial searches. However, at this stage, care was taken to examine the extent to which the participants in each study could be considered disadvantaged in their own educational context. In the discussion, consideration will be given to the extent to which this type of disadvantage is mirrored among pupils in UK schools.

The search retrieved 4786 documents, of which 1984 were found to be duplicates. Examination of the titles and abstracts allowed the following exclusion criteria to be applied:

Table 01 Exclusion criteria used in EPPI

Criteria	Number excluded
Participant age i.e. not school pupils	472
No intervention	673
Not low SES	385
Not science-related	982
Date	143

This left 147 documents to be included, based on title and abstract. However, not all of these documents were obtainable, resulting in 120 included documents. These were examined further to identify studies which reported effect sizes, yielding a list of 15 studies. Given the type of research evidence that was being sought, this would appear rather unpromising. However, the remaining 105 studies enable a rich and fascinating exploration of ways in which disadvantaged pupils may be supported in their learning of science and further participation in science education.

Appendix 6.2: Table of intervention studies with effect sizes

These were all based in the USA, except Oliver et al. (2012), which was based in Australia.

Author	Year	Number of participants	Age of participants	SES measures and levels	Intervention type	Training dose	Length of intervention	Design	Control group	Effect size
Cromley et al.	2013	1 teacher and 31 pupils (30 control)	15 years old	53% or less of mothers and 85% or less of fathers having graduated from high school	Workbook designed to support comprehension of diagrams, teacher given 2 hour PD, and then used workbooks and teacher guide in daily science lessons	1 2hr session	2 months	Quasi-experimental	Comparison	0.29 (literal comprehension (control 0.15)) and 0.52 (inferential comprehension (control 0.19))
Diaconu et al.	2012	2008-9: control: 28, treatment: 57; 2009-2010: control: 36, treatment: 60	8-10 years old	75% of pupils economically disadvantaged, urban	Rice Elementary Model Science Lab' (REMSL); one full day/week Professional Learning Communities, CPD in content and PCK for one year within model science lab environment; pre-post content tests for teachers, class observations and participant interviews	1 day per week for a year	36 weeks	longitudinal, multi-site case study	whole case sites matching demographics of test sites	0.1
Doppelt et al..	2009	control: 5 teachers,	12-14 years old	control: 74%; R, no PD:	Workshop sessions to teach reform	5 4hr sessions	2 years	multi-site case study	comparison	0.67, no treatment to

		405 pupils; reform curriculum, no PD: 5 teachers, 274 pupils; reform and PD: 13 teachers, 977 pupils		84%; R &PD: 70%	curriculum--design based learning, follow-up observations on implementation and fidelity					max, 0.5, control to reform no PD, 1.17 reform no PD to reform & PD
Finn et al.	2015	47	11 years	100%	30 minute, twice weekly physical activity intervention, data collected from activity used to teach and reinforce science concepts		6 weeks	case study	comparison	control: 0.52; treatment: 1.09
Fouche	2013	215	12-15 years old	100%	first 4 sections of introductory physics using metacognitive and self-regulatory strategies: problem solving and behaviour management; 2 45 minute sessions after each chapter exam, extra time before administering next exam;		over 6 months	Non-equivalent control group design, stratified cluster sampling by math level	comparison	0.42
Griggs et al.	2013	1561	10 years old	large ethnically and socioeconom	Responsive Classroom approach: total immersion for social and emotional		full school year	RCT	comparison	0.24

				ically diverse school district in a mid-Atlantic state'	learning; measures of anxiety and self-efficacy					
Hand et al.	2016	Teachers: yr1: 31, yr2: 32, yr3: 32, some of them the same teachers;	6-9 years old	23%; 5 school districts: 4 rural, 2 of which were rural poverty areas; 1 urban	Science Writing Heuristic approach (argument-based inquiry), 3 year study, implementation (from not implementing approach [considered control groups] to max) and analysis cycle, visited and observed in classroom by research team at least twice a year	10 days (in summer) and 3 days during year	3 years	RCT	organic	0.33
Kaldon & Zoblotsky	2014	60000 pupils, 1900 teachers, 140 district administrators and school principals	6-13 years old	70%	Leadership and Assistance for Science Education Reform (LASER) model: STC science curriculum, plus CPD	Not clear	5 years	randomized matched-pair design	comparison	0.04
Marx et al.	2004	~8000	10-13 years old	~50% (Detroit public school system)	Learning Technologies in Urban Schools (LeTUS): inquiry curriculum supported by technology: one project in 1st year, 2 in 2nd year, and 1 in 3rd year; ongoing CPD,	Monthly work sessions, summer institute and teacher support	3 years	essentially a RCT, sort of: school buy in, teachers in school selected/chosen to		measured ES from pre to post-test; by 3rd year, content improvement ES was 1.94 (10-11 years old), 0.91, 0.83 (11-12

					technology professionals in each school	groups		participate		years old), 1.39 (12-13 years old) for units.
Oliver et al.	2012	68 pupils and 6 teachers	12-14 years	57% of pupils in lowest quartile	Thinking Science Australia, or Cognitive Acceleration through Science Education (CASE): reform program,	2 full days of introduction, 6 additional CPD full days over 2 years	2 years	Case study	matched with aggregate scores reported by government	0.47
Ruby	2006	treatment: 630, control: 463	8-12 years old	school pair 1, 86/95%; school pair 2, 87/92%; school pair 3, 71/81%	reform curriculum based on NSF-supported materials, ongoing CPD, regular in-class support of teachers by expert peer coaches	Monthly professional development and in-class support	3 years	longitudinal, nonequivalent group design	matched by race and test scores	pair 1: -0.02 (non-significant); pair 2: 0.2; pair 3: 0.18
Tenenbaum et al.	2004	treatment: 30, control: 18	5 years old	72% free/reduced school lunch	children interviewed for preconceptions, drew faces to show how they felt about science (happy, sad, neutral), 2 hands on experimental lessons on bubbles, currents, and buoyancy, visited 3 exhibits at museum with hands on activities and prediction/argumentation, re-interviewed for content knowledge	1 session to plan the lessons	1month	multi-site case study	comparison	buoyancy: 0.11; bubbles: 0.24

Thomas et al..	2015	treatment: 214 and 295 (subsequent years), control (pre-matched from NYC) 19,392 and 27,548	13-15 years old	78% free/reduced school lunch overall in NYC, 64-92% in participating schools	Peer-Enabled Restructured Classroom (PERC): grade-adjacent peers lead small groups, ask questions, scaffold thinking, encourage, formatively assess performance	Peer leaders – extra class per week, teachers – not clear	full school year	multi-site case study	comparison	odds ratio on passing/failing: 1.46 (only in 2nd year).
Tong et al.	2014	Pupils: 94 treatment, 194 control; teachers: 5 treatment, 7 control; Schools: 2 treatment, 2 control.	11 years old	85% free/reduced school lunch	structured and scripted lesson plans w Inquiry-Based Learning, vocab, reading & writing integration into science; daily writing prompts, science content reading, and pupil recorded glossary in science notebooks; (except for physics)	ongoing twice a week CPD plus twice a month meetings and classroom observations (for fidelity)	school year	RCT	comparison	odds ratio on passing/failing: 10.28, treatment groups outperform control groups on standardized state tests due to literacy *integration*

Appendix 6.3: Recommendations for successful science interventions

From Bell (2014)

Recommendation 1: Initiators, developers and other stakeholders should ensure that interventions have a clear purpose meeting well-defined needs to address and overcome a problem which is well-evidenced and articulated.

Recommendation 2: Despite the progress that has been made in recent years, greater efforts are still required by all parties to bridge the communication gap between teachers and originators of interventions both big and small.

Recommendation 3: All parties involved in interventions should give a higher priority to the use of existing evidence to inform the design of interventions and to the collection and use of evidence as an integral part of the intervention. There should be: clearer reasons for gathering evidence; a better match between the type of evidence collected and the questions that are being addressed; and a strengthening of the processes for monitoring progress and impact of the intervention, including unexpected outcomes.

Recommendation 4: Further efforts are needed to improve the evaluation of interventions in order to strengthen the contribution it can make to the outputs and outcomes of interventions. This could involve improved guidelines from funders, training for practitioners involved in interventions, and reviews of families of evaluations to consolidate findings on the effectiveness of the interventions and on the process of the evaluation itself.

Recommendation 5: Further consideration needs to be given to:

- additional research to understand better how interventions can be applied effectively to new contexts
- greater emphasis on support and training for implementing the intervention when it is introduced into a new context.

Recommendation 6: The landscape of interventions does not get any less complex with time, therefore all stakeholders – including policy makers, funders, researchers and practitioners – must increase their efforts to engage in open dialogue on interventions in order to establish need, effectiveness, quality and value for money. Particular consideration should be given to:

- revisiting ways to rationalise the number of interventions in science education, increasing the number of collaborative programmes
- developing an 'intervention toolkit', similar to that published by Education Endowment Foundation, specific to science education and designed to inform practitioners of the range in interventions available, the evidence base for their effectiveness and value for money.

Recommendation 7: Greater emphasis must be given to ensuring that implementation of interventions is to the highest possible standard. In particular, more effort should to be put into supporting schools and practitioners to ensure they:

- are party to the development of the intervention
- have the necessary expertise, skills and knowledge to make informed judgements on which interventions to choose, and how to implement and evaluate them by making better use of existing research and their own evidence and experience

- are engaged in relevant professional development for continuous improvement in their practice.

Recommendation 8: Further research should be undertaken to understand better the processes which contribute to successful interventions, in particular, those which bring about effective and sustainable change in the behaviour of individuals and organisations.

Recommendation 9: Consideration should be given to testing and refining such a model for developing interventions in order to explore in more depth ways in which interventions of all types can be made more successful.