Education research

The perceived success of interventions in science education

May 2014

wellcome trust

A report for the Wellcome Trust

by

Derek Bell
Campanula Consulting

Acknowledgements

This report would not have been possible without the co-operation of those individuals who willing gave of their time to take part in the interviews and focus groups and share their views, and the support of staff in Medical Humanities and Engagement division of the Wellcome Trust. The help and contributions of everyone involved are very much appreciated and acknowledged here.

Contact details

Any correspondence regarding this report should be sent to:

Professor Derek Bell
Campanula Consulting
4 Rowan Park
Christleton, Chester CH3 7AZ
Email: derek@campanulaconsulting.co.uk
Telephone: 01244 336207
Mobile: 07852 436277

Contents

Executive summary

Over many years much effort and funding has gone into developing interventions to improve standards in education. Although the majority of interventions lack high quality evidence about their effectiveness, stakeholders, including policy makers, researchers, teachers and subject communities, have still made judgements about their levels of success. Critically, decisions are being made throughout the system about which interventions to adopt, promote, fund, continue or stop. This raises the question: *In the absence of robust evidence, on what grounds are interventions considered to be successful or not?*

The purpose of this study, therefore, was to explore, with specific reference to science education, what a range of stakeholders consider to be the features of successful interventions. This was designed as a qualitative study and adopted an iterative approach which involved a search of existing literature, semi-structured interviews with a total of 30 individuals from range of backgrounds, and focus groups with 15 teachers, all based in England.

A fully comprehensive review of the literature was beyond the scope of the current study, but by using exemplars, three key elements were identified as contributing to successful interventions: the underpinning principles, the expertise of the personnel and the context of the intervention.

Participants in the study emphasised that judging the success of any activity in any sphere involves considering a number of factors. Overall, interviewees were positive about interventions in science education, acknowledging that there were very few in their experience that had nothing to offer to improve young people's experiences of science education. The interventions identified during the study could be divided into five groups according to their underlying stimulus: policy-driven, knowledge-building, pedagogical improvement, curriculum development, or enrichment and enhancement.

Throughout the study, participants reiterated that no two interventions are alike nor do they follow the same pattern in terms of their development or impact. There was a considerable degree of consensus around the criteria that might be used to judge success; these fell into six categories: levels of take-up, improvements in attainment, engagement of pupils, changes in practice, changing the terms of the debate, and value for money.

Despite the emphasis placed on the need for robust research and good evidence to develop successful interventions, the perception is that in practice little use is made of either existing evidence or that which is gathered during the intervention itself.

There was a consensus that successful interventions depended on a combination of key elements which included:

- a clear definition of the purpose of and need for the intervention
- the clarity of the process
- the effectiveness of the implementation
- the suitability of the people involved
- the appropriate level of monitoring, evidence and accountability
- the quality of communication, promotion and profile it achieves.

Recommendations

Reflection on the views expressed identifies seven cross-cutting issues, leading to nine recommendations.

The clarity of purpose and shared understanding of a successful intervention

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.*

The quality, quantity and nature of evidence used to define and judge success

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.*

The degree to which the situational context affects the likely success of an intervention

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.*

The extent to which the impact of policy changes might hamper or support the initiation and development of successful interventions

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.*

The challenge of implementing interventions successfully

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, implementing and evaluating 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.*

The extent to which effective change management might contribute to successful interventions

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.*

The need for a more holistic model for developing interventions

A model is proposed in an attempt to bring together the lessons derived from the discussions and to form the basis for developing successful interventions. The different elements all interact and form three intersecting axes in which:

- the clarity of, and commitment to, the **purpose** lead to tangible impact and **outcomes**
- suitable **people** working in the right **context** results in measureable and demonstrable **outputs**
- robust **processes** lead to effective **implementation**.

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.*

1. Introduction

For many years, much effort and funding has gone into developing initiatives to improve standards in education and consequently the achievements of young people. As measures of performance, both internationally and nationally, have increased, comparisons between countries and between schools have become easier and the findings more accessible. As a result many governments and other sectors of society have taken steps to improve educational standards and address identified shortcomings. In broad terms such interventions have been brought in at:

- national level, to raise standards across the country (e.g. National Strategies in England)
- local level, to address issues in particular areas (e.g. City Challenge programmes targeting schools in inner cities in England)
- school level, to improve overall performance to meet internal or external targets.

Without question there are a great many initiatives but assessment of their impact and success is complex. There is variation not only in the scale and type of initiative but also in the criteria against which success is judged. In particular, there appears to be a lack of high-quality evidence about the effectiveness of the vast majority of initiatives. Furthermore, given the large number of initiatives, knowledge of them is often restricted to particular geographical locations, sectors, communities or groups of individuals.

Despite this lack of knowledge and evidence, many initiatives are perceived to be successful (or not) by stakeholders including policy makers, researchers, teachers and subject communities. Critically, decisions are being made at all levels of the system about which interventions to adopt, promote, fund, continue or stop. This raises the question:

In the absence of robust evidence, on what grounds are interventions considered to be successful or not?

The purpose of this study, therefore, was to explore, with specific reference to science education, what a range of stakeholders consider to be the features of successful interventions.

Context of the present study

In the UK, as in many other countries, particular concerns have been expressed with regard to science education because of the important link that is made between it and the performance of the economy. The predicted demand for science skills alongside technology, engineering and maths (collectively referred to as STEM subjects) in the workplace over the next five to ten years[1] has provided a major stimulus for a large number of interventions in this area. For example, the National Audit Office (NAO, 2010) reports that in 2004 some 478[2] interventions designed to improve young peoples' experiences of STEM were identified as part of a review by the Department for Education and Skills (now the Department for Education). Despite proposals in the 2006 STEM Programme Report (Department for Education and Skills, 2006) to rationalise the number of interventions and reduce overlap, a large number remain.

[1] See for example Harrison (2012).

[2] Of these, 120 were supported by Department for Education and Skills, 217 by other government departments and 141 by other organisations.

This study looked at science-focused interventions but inevitably there is overlap with those related to other STEM subjects and to teaching and learning in general. This report should be read in the context of wider influences in education policy in England, including the previously highlighted drive to improve education and the importance of science to the economy. Other contextual factors include:

- greater emphasis on value for money
- increased accountability and transparency
- calls for evidence-based policy and practice
- the drive to increase the autonomy of schools (e.g. through expansion of the academies programme, the introduction of free schools, studio schools and University Technical Colleges) and greater independence in relation to the whole curriculum, teaching, learning and financial control.
- the reduction of local authority support, so that teachers need to look elsewhere for it.

These issues highlight the importance of making good decisions on which interventions to adopt and understanding how these decisions are made as well as improving the likelihood of an intervention being successful.

2. Methodology

This was a qualitative study designed to explore four questions:

- Which interventions to improve the quality of teaching and learning in science education are considered to be (or have been) successful by a range of stakeholders and why?
- What are the factors that are considered to contribute to the levels of success of interventions to improve the quality of teaching and learning in science?
- To what extent do the perceptions of stakeholders reflect the available evidence as to the levels of success of interventions?
- In what ways might the effectiveness and impact of interventions be improved in order to establish sustained improvements in science education?

The study, which adopted an iterative approach, was developed in four phases.

Phase 1 and ongoing: Desk research

Searches of existing literature on what defines a successful intervention were undertaken. The principal sources used include internet search engines (Google and Google Scholar), British Educational Index (BEI), Education Research Index Catalogue (ERIC) and discussions with representatives from organisations (e.g. the Centre for the Use of Research and Evidence in Education[3] and the Institute for Effective Education[4]) promoting evidence-based education.

Phase 2: Preliminary semi-structured conversations

Ten semi-structured exploratory conversations, approximately 45-60 minutes long, were undertaken either by telephone or face-to-face with individuals drawn from a range of stakeholders including science education researchers, teachers and policy makers. The purpose of these preliminary conversations was to help scope the study and clarify the problem and definitions of the principal terms to be used and the issues to be addressed during the main phase of the study, including:

- examples of what might be considered successful interventions
- factors that contribute to the success of interventions
- criteria by which success might be judged
- to what extent the perceptions of success are supported by research evidence
- challenges involved in extending the influence of interventions so they are adopted more widely and made sustainable.

Phase 3: Interviews with a range of stakeholders

This phase involved semi-structured interviews with an additional 20 individuals including government officers, researchers and curriculum developers, representatives of professional bodies and learned societies, funders, teacher educators, local authority advisers/consultants, and science communicators. The experience of the interviewees also covered primary and secondary phases of education as well as non-school-based programmes.

---

[3] See CUREE, 2011.

[4] At the University of York, UK. See york.ac.uk/iee (accessed 3 April 2014).

Prior to the interview, participants were sent an outline of its purpose and a table that contained a list of interventions covered in phase 2 (see Appendix 1). They were asked to add further interventions to the list and to indicate whether they considered each intervention to be successful or unsuccessful. Thus interviewees had time to give some thought to the issues and possible interventions before the interview.

The interviews, which were conducted February to April 2012, followed a set format (see Appendix 2) in which interviewees were asked to:

- reflect on their responses to the selection of interventions provided before the interview and the additional examples they identified, giving reasons for their choices
- indicate any research evidence they were aware of to support their decisions
- outline the criteria they would use to judge the success of an intervention
- suggest what elements are needed to ensure that an intervention is successful
- set out the challenges they saw in implementing interventions and getting them adopted more widely and sustainably.

Phase 4: Teacher focus groups

Two focus groups were held in July 2012 (one with ten teachers and one with five) from two secondary schools (one in the north-west and one in the north-east of England). The focus groups followed the same structure as the interviews, with the teachers being sent the notes on the project and the list of interventions for completion beforehand. The responses to the list were used as the basis for the first part of the focus group and the interview template was used to structure the remainder of the discussions.

3. Existing literature

The education research literature contains many reports of studies which have examined the effects of interventions to improve teaching and learning. The majority have been small-scale, often undertaken in only one or two schools and lasting for a short time: weeks and months rather than years. Despite the range of material there is very little that has specifically analysed what it is that makes some interventions judged to be successful and others less so. A fully comprehensive review of the literature is beyond the scope of the current study but this section provides examples of four types of publication which exist and draws out the issues that need to be addressed in attempting to identify what is perceived to make an intervention successful.

System-level reports

These reports aim to analyse the effects of large-scale interventions which have usually been initiated by government. The following three examples are of particular interest to the current study and have been included to illustrate different perspectives on how the question of what makes interventions successful might be addressed.

The first, which was not science-specific (Mourshed et al., 2010), reports on an analysis of 20 school systems from around the world, from 1995 to 2007, that have succeeded where others have failed to make improvements. In particular the study endeavoured "to understand precisely which interventions occurred in each school system and when and how these interventions interacted with each other and with the system's broader context to deliver better outcomes for students" (p. 7).

Successful, improving systems were found to demonstrate a striking consistency in the interventions they employed as they progressed through a three-step process of:

- establishing the current levels of performance in order to clearly identify the starting point identifying the specific set of interventions needed to make the desired improvements in student outcomes
- adapting the intervention cluster to the prevailing context taking into account history and culture.

Two additional ingredients for success were identified:

- understanding the stimulus for bringing about the necessary changes
- sustaining the improvement over the longer term through features such as developing strong pedagogy supported by collaborative practices and continuity of leadership.

The report notes that "improving systems appear to be careful in maintaining the integrity of the interventions" (p. 26) and some interventions are more appropriate to the particular stage of development of the system in question. For example, moving from "poor" to "fair" tended to require strong central guidance with a focus on numeracy and literacy, but the core driver to move from "great" to "excellent" was peer-led creativity and innovation. In addition, there were a small number of cross-stage interventions – for example, curriculum and standards revision, improving student assessment and building the skills of teachers and head teachers – that contribute to improvement but the effects of which differ according to the stage in which they are applied.

Despite the wide variation and complexity of the systems studied, it was clear that selecting the most appropriate interventions was the highest priority, while setting these in context is also important. Increasing spending on interventions does not automatically lead to increased performance; successful systems achieved better value for the money that was available.

The second report (NAO, 2010), in which the issue of value for money is addressed directly, provides a detailed analysis of STEM-specific interventions in England between 2004 and 2010. During this period, significant emphasis had been placed on increasing the number of young people continuing science beyond the age of 16 and going into science or science-related careers. It was estimated that an additional £40 million was spent on activities targeted at young people up to the age of 18 specifically to address the matter.

The NAO evaluated the "trends in take-up and achievement in science and maths", and assessed "the effectiveness of major programmes in supporting progress". Five critical success factors were identified as key to improving take-up and achievement in science and maths:

- careers information and guidance
- quality and quantity of school science facilities
- quality and quantity of science teachers
- image and interest
- availability of separate GCSE sciences ('Triple Science').

The analysis was, however, severely hampered by the fact that of the 478 interventions that were considered, two-thirds had no evaluation or had no evaluation planned. The use of multiple regression techniques enabled researchers to argue that three of the intervention types were associated with statistically significant increases in numbers of pupils achieving A*-C grades at GCSE in sciences. The interventions were: enhancement and enrichment activities, continuing professional development training at the National Science Learning Centre, and STEM Ambassador activities.

The report concludes: "Our analysis suggests that the interventions [examined] are associated with improved take-up and achievement in science and maths, but that they could be rationalised and provided to schools in a more systematic way" (p. 42). The difficulties of separating out the effects of individual interventions and the crossover effects of multiple interventions are noted. The authors also caution that those interventions which are not associated with statistically significant changes in achievement are not necessarily ineffective: "The incompleteness of activity data… as well as the relatively short timeframes and difficulty of establishing causal links to take-up and achievement, mean that more evaluation is required to conclude on longer-term effectiveness" (p. 42).

The third example is the most recent (Higgins et al., 2011; updated in Higgins et al., 2013); it was commissioned by the Education Endowment Foundation[5] to support schools in implementing a major policy initiative in England – the Pupil Premium. Described as part of a 'toolkit' to support teaching and learning across the curriculum, the report sets out a series of interventions, some of the research evidence for them (and the strength of that evidence), their impact, value for money, and which phase of education and core subject (English, maths, science) they apply to. It emphasises that some of the most effective interventions do not require large sums of money.

Besides providing a compendium of useful information, one aim of the toolkit is to encourage schools and teachers to make informed choices through better use of research evidence. The toolkit also emphasises that it is up to schools to identify approaches that meet their needs and to develop, monitor and evaluate them. It is early days but a report on the way schools use Pupil Premium funding

[5] For more information see educationendowmentfoundation.org.uk (accessed 3 April 2014).

(Carpenter, 2013) indicates the tensions between the use of academic research evidence and personal experience in making decisions on interventions. This point is referred to again in Section 6.

Systematic reviews of generic issues

There are a variety of systematic reviews – which vary according to the criteria used for including studies in their analyses. For example, a meta-analysis (Schroeder et al., 2007) of the effects of teaching strategies on student achievement in the USA identified eight categories of teaching strategies which show significant effects. The four with the strongest effects were enhanced context strategies (effect size 1.48), collaborative learning strategies (0.95), questioning strategies (0.74), and inquiry strategies (0.65). These findings resonate with a synthesis of evidence conducted by CUREE on curriculum change in England 2007 and 2010[6]. This analysis focused on classroom implementation and found that there were positive outcomes when the changes involved experiences that:

- placed ideas, facts and phenomena in context
- built on students' existing knowledge, understanding and skills and engaged learners actively in assessment
- promoted conceptual development and encouraged cross-curricular transfer of learning
- involved structured group work and effective talk as a means for students to access the curriculum
- included curriculum tasks that were specifically planned to challenge all pupils
- were underpinned by excellent subject knowledge and professional development of teachers.

Systematic reviews of interventions specifically related to science education

Examples of systematic reviews specific to science are relatively rare but, based on the findings of studies undertaken by the EPPI Centre[7], a series were published between 2003 and 2006 (Bennett et al., 2003; Bennett et al., 2004; Lubben et al., 2005; Hogarth et al., 2006). Collectively they found reasonable evidence to support claims that include:

- context-based approaches foster more positive attitudes to science, motivate pupils and do not adversely affect pupils' understanding of scientific ideas
- small group discussions have beneficial effects on learning but teachers and students need to be given explicit instruction (e.g. in developing arguments)
- students' use of ICT simulations improves basic science ideas but for higher levels of understanding there is little or no difference.

More recently, Slavin (2012) published a review looking specifically at a synthesis of programmes in primary (elementary) school science. This emphasised the lack of studies that met the criteria for inclusion in the review: "of 327 identified studies purporting to evaluate science approaches in elementary schools, only 17 had control groups, durations of at least four weeks, equivalence on pre-

[6] For summary, see curee.co.uk/files/publication/[site-timestamp]/Curriculum%20Session_0.pdf (accessed 3 April 2014).

[7] Further information on the EPPI Centre can be found at eppi.ioe.ac.uk/cms/ (accessed 3 April 2014).

tests, and measures not inherent to the experimental treatment". Consequently the findings from the analysis have to be treated with caution but the three main ones are worth noting.

The first and unexpected finding came from "the largest and best-designed of the studies", which concluded that elementary science programmes that provide teachers with kits to help them use regular hands-on, inquiry-oriented activities had limited impact on achievement. The second finding was that studies of inquiry-oriented professional development programmes, which did not provide kits, demonstrated positive outcomes in science achievement. The third finding was that the use of technology has potential, although there was a need for "further development and large-scale evaluations of modern approaches that integrate video and computer technologies with inquiry-oriented teaching and cooperative learning".

Reports of specific intervention projects related to science education

Reports on specific interventions related to science are comparatively few; CASE (Cognitive Acceleration in Science Education; Adey et al., 1989), based on the explicit development of thinking skills, probably has the strongest research base. Interestingly, there was no immediate effect on achievement in science as measured by the end-of-year science examination but a year later the CASE pupils performed significantly better in science than the control group (Adey & Shayer, 1990). Two years after the intervention, when the pupils who had begun CASE in Year 8 took their GCSEs, boys' performance in science, maths and English was significantly better in the experimental group than the control, although for girls the difference was only evident in English (Shayer & Adey, 1992). Such findings have not gone unchallenged (Jones & Gott, 1998) but evidence to support CASE remains substantial.

A very different type of study (Springate, 2009) examined the impact of the STEM Pathfinder programme which was piloted by 40 schools between October 2008 and June 2009. Set up as an evaluation study, as opposed to a piece of academic research, this report explicitly identified, through self-report by the personnel involved, key characteristics of successful STEM Pathfinder activities:

- there was an individual or group responsible for overseeing STEM activities supported by senior leadership teams with sufficient time for teachers to meet and plan collaboratively
- activities were delivered by enthusiastic teachers willing to try something new, and involved external partners (principally from industry)
- activities had a clear focus, a 'real-life' context, a competitive element, some freedom for students to experiment and think for themselves, practical and interactive aspects, and a good balance across all STEM subjects.

Themes from the literature

From this short review of literature it is possible to draw out some tentative themes as to what contributes to interventions having positive effects. Although the details vary, the majority of the themes appear to apply to interventions at all levels. Three elements might be proposed:

Underpinning principles: ***Reflecting a key finding of the Mourshed et al. (2010) analysis, the success of interventions is related to their clarity of purpose. More specifically, successful interventions are designed to meet a defined need or overcome a***

particular challenge, share fundamental strategies which can be adapted for different contexts, and require a combination of approaches to achieve their goals.

Expertise of the personnel: ***Successful interventions depend on the personnel involved and the quality of their expertise. In particular they depend on teams of people including individuals with a range of leadership and management skills and teachers with high levels of pedagogical skills and subject knowledge (maintained through professional development), and*** collaborative partnerships between organisations and communities.

Context for the intervention: Interventions do not take place in isolation – context, culture and history can have a major influence on their success. Initial considerations include: the starting point of the system (the school, the teachers, students and other personnel and organisations); how to maximise the features of the situation which are likely to enable success (and minimise those that militate against it); and funding (although simply increasing the money available will not guarantee success).

These themes recur in the sections that follow, informing the analysis and discussion of the views expressed by the interviewees and members of the focus groups.

4. Defining terms: ‘intervention’ and ‘successful’

Prior to the interviews and focus groups the following information was included in the details sent to participants (see Appendix 1):

> **Intervention** is used to refer to programmes and activities that aim to improve teaching and learning in order to raise achievement and improve learning experiences in science education. It includes activities designed to alter approaches to teaching and ways in which the curriculum is implemented. For the purposes of this study therefore, the term **intervention** will be used generically to cover a wide range of actions and activities. Where it refers to a particular type of intervention, e.g. a specific teaching technique such as questioning, this will be made clear.
>
> **Successful** will be used as an all-embracing term to describe the achievements of an intervention and can include a wide range of criteria. The actual criteria or measures used are likely to vary according to the role and position of the interviewee. Indeed, the way in which individuals describe the success of interventions is part of the study. **Effective** will be specifically restricted to the assessment of an intervention against its stated objectives.

This section explores the interpretation of these terms in the light of the responses from the interviews and the focus groups.

What is success?

Ultimately, judging the success of any activity in any sphere involves considering a number of factors. In some circumstances success can be measured against specific criteria, say, clearing the bar in a high jump competition. Yet, even this might be open to some degree of interpretation if, as a coach, you are also looking for some improvement in technique. Almost inevitably success is never ‘black and white’ and the more complex the activity being undertaken, the more this is true. All the contributors to this study were very aware of the difficulty and were at pains to justify their own particular stance in responding to questions. Some were almost apologetic as they explained that their “views are unscientific decisions and judgements” [I08], “are not necessarily research-based” [I09] and that “personal experiences were very influential” [I12] when looking at the examples of interventions. Four particular caveats were commonly expressed in some form.

The first was that success can vary according to the level at which it is judged. As one interviewee observed, “reported success does not necessarily reflect the reality on the ground” [I09]. This is not to imply any manipulation of information (although it was suggested that this is not unknown) but to highlight that focusing on particular criteria for reporting can lead to a misleading picture of the success of an activity, project or programme. For example, some interviewees referred to the challenge of meeting the requirements of a funder as opposed to the learning gains that might be achieved by pupils: both are important but they can conflict.

The second caveat is that it is possible for a project, activity or organisation to be successful when assessed against some criteria but unsuccessful against others. One interviewee illustrated this by

reference to an ex-Premier League football club, which he described as being reasonably successful "on the field", having won a major trophy in recent years, "but financially it is a disaster" [I07], having gone into administration. It was suggested that an example of such a situation in science education was the QCA science schemes of work for KS1 and KS2[8]. These were perceived by at least one interviewee as being successful in providing a helpful structure for the curriculum in some primary schools, but unsuccessful in that too many schools became reliant on them, which constrained further improvement.

The third caveat was stated as a warning – "Don't confuse the success of an activity with the personality of the people involved" – and as a question – "Would this project continue if it wasn't for [name of an individual]?" [I19] Such overdependence on a particular individual may mean that a project lacked underpinning structures which could prevent its demise when that person moved on.

The fourth caveat was the influence of time and timing. Success may depend on the time elapsed since an intervention was introduced. The timing of an intervention can also play a major part in determining how successful it is. One interviewee referred to them as "heroic failures" [IC01], explaining that the projects seemed to have all the ingredients for success but at the time of their launch there was a change in curriculum focus which mitigated against their adoption.

Section 5 discusses perceptions of specific interventions and Section 6 examines in detail the criteria against which success of interventions might be judged. What follows here are the common features that framed how interviewees expressed judgements of success.

- *The extent to which an intervention met its objectives.* This was the starting point for the majority. As one interviewee noted, the key question was "Did it achieve what it set out to achieve?" [I02]
- *The effect on the audience or beneficiaries.* The reaction of teachers was considered to be a significant indicator. "Do teachers talk about it?" [I07] and "Did the intervention change the practice of the teachers and what they teach?" [I05] were typical questions. "Teachers must see an immediate need for the intervention and that it is going to help their teaching" was a further sentiment frequently expressed. This was often linked to "How do children react?" [I06]. Occasional reference was also made to the way in which other stakeholders and the media present the intervention.
- *The principles on which the intervention is based.* This consideration was presented in different ways that, to a large extent, reflected the "Personal sphere of knowledge" [I03] of the individual making the judgement. For some, the underlying theories of learning were a significant element and for others the issue was more pragmatic, as in "Why are we moving in the direction determined by the intervention?" [I05]. Overall there was a common concern to know why an intervention had been developed and whether or not it concurred with the wider considerations such as personal philosophy and attitudes of those implementing it.
- *The wider influence of the intervention.* It was generally acknowledged that reach and influence of interventions varies greatly. Some are local and small-scale and may be high-quality but little-known. Others are largely knowledge-building or experimental, so they may have limited initial impact but the potential to have greater influence over time. On the other hand, larger-scale

[8] For further details, see webarchive.nationalarchives.gov.uk/20090608182316/http://standards.dfes.gov.uk/schemes3/subjects/ (accessed 3 April 2014).

interventions may be widely known and have potential for greater influence on short-, medium- and long-term attitudes and values as well as the quality of teaching and learning.

- *The return on investment.* This involved not simply the cost in money but also the level of commitment expected from those involved in the intervention and its implementation.
- *The type, quality and quantity of evidence available.* This consideration was expressed both in terms of the research evidence available and the measures that are used. In both cases it was acknowledged that the evidence is fragmentary and not necessarily what is needed. As one interviewee expressed it: "We don't have good measures of success and are very reliant on short-term measures… there is little or no joint recording of evidence… bringing together different forms of information to provide a more holistic picture of the intervention, its outputs, outcomes and impact" [I02].
- *The multi-dimensional nature of the problem.* Determining whether or not an intervention is successful is a multi-dimensional issue. "No single intervention is unproblematic in its success" [I05]. As indicated in the four caveats presented above, all interventions have limitations, and judgements about their success are rarely made without reservations.

What is an intervention?

Although there was a general acceptance that, in this study, the term 'intervention' could be applied to actions that are designed to improve the quality of teaching and learning in science, the discussions revealed a wide range of interpretations. Further terms – 'initiative', 'programme' and 'project' – were frequently used. Broadly, these were applied as follows:

- 'Initiative' was most commonly linked to policy developments, usually instigated by government, but it was also applied to the introduction of changes in policy at the school or science department level.
- 'Programme' and 'project' were used almost interchangeably to refer to a range of activities of differing scale and size that involved a series of actions carried out over a period of time either in sequence or in parallel. These might range from the introduction of a new scheme of work in school to a national curriculum project or a programme of enrichment and enhancement activities.
- 'Intervention' tended to be reserved for pedagogical approaches and more clearly defined activities which were directed at bringing about a direct change within the classroom and the learning of a particular group of pupils.

Despite having provided all the interviewees with the same information prior to their participation, the difference in the approach adopted between those who were interviewed individually and the teachers who made up the focus groups was striking. The teachers consistently used the term 'intervention' in a much more restricted sense, referring most frequently to specific actions that they undertook in the classroom or that were being introduced in their school.

Although it was not reflected specifically in the terms used, there was often a thread in the discussions which referred to the target audience for the intervention. For the majority of interviewees there was an implicit assumption that the intervention was directed mainly at teachers, who would be the channel by which it would be enacted and brought to pupils. Other audiences explicitly identified were school leaders and the pupils themselves.

Effecting change was a strong thread running through all conversations. A core purpose of interventions was seen to be attempting to bring about change; how well each did that was a mark of its success or failure. This is returned to in more detail in Section 7.

In this report, the term ‘intervention’ will be used as defined at the start of this Section (p. 15) to include the wide range of actions and activities outlined in the previous paragraph. When it is used to refer to a particular type of intervention, this will be made explicit.

5. The interventions

This section describes the interventions that have been identified during the current study. In contrast Section 6 specifically addresses the criteria, as identified by the interviewees and focus groups, by which an intervention might be considered successful or not.

Overall interviewees were positive about interventions in science education, acknowledging that there were very few in their experience that had nothing to offer to improve young people's experiences of science education. This attitude could in part have been the context in which the discussions were set, i.e. exploring 'perceptions of successful interventions', but it did not prevent the expression of strong views about some interventions. For example, the proposed introduction of a Science Diploma was described as "catastrophic" and "a clear example of a need for market research to find out if it is really needed before doing it" [IC01].

None of the interventions described as successful were seen to be without shortcomings. The SATIS project, for example, was widely considered to be successful because it "changed perceptions and approach to science but the language was high level so it was not accessible for all students" [I10]. Particular concerns were expressed about the unintended consequences that resulted when interventions become implemented in a way that was at odds with the original purpose. Attempts to introduce assessment for learning, for example, were referred to by some interviewees in this light: "At one level this is successful but not really translated into practice – it's been taken over by policy." [I2] and "QCA [Qualifications and Curriculum Authority] implementation is not successful, APP [Assessment of Pupil Performance] is unconvincing" [I15].

Identification of the interventions

All the interventions that were identified by the interviewees are listed in Appendix 3. It must be emphasised that there has been no explicit or implicit attempt in this report to label any of the interventions as successful or unsuccessful. References to specific interventions are presented as illustrations of perceptions of success and not as judgements on the interventions themselves. The wide range of interventions referred to by interviewees reflects the diversity of backgrounds of the individuals involved.

A typography of interventions

As shown in Appendix 3 it is possible to place the interventions identified during the study into five groups according to their underlying stimulus – policy, knowledge-building, pedagogical improvement, curriculum development or enhancement and enrichment activities. The boundaries of these groups are not rigid nor are any of the interventions restricted to a single area. Taking this approach further demonstrates how the criteria for success vary according to, among other things, the underlying purpose of the interventions in question.

Policy-driven interventions

Policy-driven interventions endeavour to create an environment in which new approaches can be introduced and successfully implemented in order to bring about the change required.

The examples of national policy-driven interventions that were referred to were directed at bringing about system-wide change. As such, success was generally defined in terms of the numbers of schools that had adopted it (e.g. Triple Science), increases in examination performance (e.g. National

Strategies) or the number of students continuing to study science beyond 16. Set against these headline criteria other factors either strengthened the claim for success of detracted from it. For example, the contribution of the National Strategies to raising examination performance was strongly contested by some interviewees when the cost of them was taken into account. Consideration of the Strategies also exemplified very clearly the way in which the perceptions of interventions can change over time: the Strategies were, in the early stages, judged effective in tackling some pedagogical issues, but in the later stages over-complex and expensive.

It is worth noting that only one interviewee mentioned what might be considered the biggest policy-driven intervention in the last 30 years, the National Curriculum for England, Wales and Northern Ireland. It can be argued that the arrangements for curriculum and assessment that this enshrined in legislation have overshadowed every intervention since 1989. Although there was no discussion during the current study of the perceived success or failure of the National Curriculum *per se*, its requirements were nonetheless taken to be the backdrop against which many of the curriculum interventions were judged. Indeed, the origin of many interventions can be traced back to efforts to make the National Curriculum more relevant to young people – the development of Twenty-first Century Science being a particular example.

Accountability and inspection measures also influence the way in which interventions are perceived. This was particularly evident in the discussions with teachers. The pressure felt by schools and teachers to meet the government performance measures and the demands of Ofsted inspections means that they cannot ignore policy-driven interventions. For some schools, reaction to policy announcements is a high priority. At one extreme one interviewee, who works in a policy role, suggested that the announcement in 2010 of the English Baccalaureate (EBacc) was an extremely successful intervention. As a result, schools, almost immediately, began to alter the balance of subjects available to students in KS4 in order to meet the new EBacc requirements. The argument put forward was that the announcement brought about a significant shift in behaviour in line with government policy at no direct cost to the Department for Education.

Although policy-driven interventions are mostly discussed in terms of national policy, some interventions are adopted because of policy changes in individual schools. For example, the Primary Science Quality Mark (PSQM) can be considered to be an intervention, indicative of a school having taken a policy decision to raise the quality and standards in science. Undertaking PSQM requires a whole-school approach and is unlikely to be successful if it has not been taken on across the school.

Knowledge-building interventions

Knowledge-building interventions develop new understandings and thinking about ways of improving teaching and learning, and provide evidence of the effectiveness of interventions that, in turn, generate further knowledge and inform new thinking and understanding.

It can be argued that all interventions are and should be knowledge-building but this relatively small group of interventions were identified because they were perceived to be successful in providing information to teachers and to have had influence well beyond the initial projects. The research base was considered an important factor in the perceived success of three of the interventions – CASE, SPACE Project and CLIS[9] – but each also provided extensions which translated the research for the

[9] The Children's Learning in Science Project. More details can be found at: nationalstemcentre.org.uk/elibrary/collection/464/children-s-learning-in-science-project (accessed 3 April 2014.)

classroom. In addition, each of the interventions was viewed to be successful because of the wider influence it has had over a substantial period of time and on other developments in science education.

Pedagogical interventions

Pedagogical interventions provide teaching and learning strategies that are used to engage students in learning in order to improve their skills, knowledge and conceptual understanding, both generally and in specific disciplines.

Examples of pedagogical interventions ranged from small changes in the way a teacher uses questions to larger interventions that become adopted as policy. The former were uppermost in the mind of many of the teachers, who referred, for example, to the need for "wait-time" when asking pupils questions, and the use of resources such as "concept cartoons" and "badger tasks", among other things. These interventions were felt to be successful when they helped to engage pupils with starting points in lessons or revision of ideas and topics. Concern was expressed by several interviewees that interventions that rely on specific resources can fade after the initial enthusiasm, with the actual pedagogy slowly reverting to previous practice. The effect of the intervention is short-lived because it has not become fully embedded in the way teachers and pupils work.

The success of larger interventions was similarly considered variable for a variety of reasons. The most frequently mentioned problem was the way in which an intervention can become overtaken by events so that its initial impact becomes diminished. The effect of the National Strategies has already been referred to (p. 21) and the introduction of Assessment for Learning (Afl) was mentioned by several of the interviewees. For the majority, Afl as an idea was considered to be a successful intervention, given impetus by the publication of *Inside the Black Box* (Black and Wiliam, 1998b). This pamphlet, based on an extensive review of existing research (Black and Wiliam, 1998a), provided strong evidence that pupil attainment can be raised significantly by adopting pedagogical approaches that involved formative assessment. However, as one interviewee explained, "At a rhetorical level this [Afl] is very successful but it has not really been translated into practice…it has been taken over by policy" [I12] – a thought reflected by another interviewee, saying "We can talk about this very fluently, but don't really know what it is in practice" [I20].

The lack of success in translating Afl, or more accurately formative assessment, into practice is because it is difficult to implement and depends on extensive support and training for teachers. However, that aside, there were many concerns expressed that the lack of subsequent success of this and other interventions was in part because they were taken over (some people would say "hijacked") and implemented in ways that conflicted with the original purpose. In contrast, it was argued that an intervention which retains its focus is more likely to be considered successful. Several interviewees felt this was one of the strengths of the Stimulating Physics Project, which was targeted and had the added feature that it worked directly with teachers.

Curriculum interventions

Curriculum interventions explore ways in which particular skills (academic, practical and interpersonal), knowledge and conceptual understanding can be provided for students in contexts which help generate interest in, and enthusiasm for, learning. Curriculum interventions were referred to most frequently. The majority of examples given pre-date the National Curriculum, the implication being that since its introduction, the scope for innovation in curriculum development has been more constrained. Although they no longer exist in any formal sense, some of these interventions were

clearly felt to have introduced an alternative approach to science education and to have influenced subsequent developments.

Reflection on one of the most recent major curriculum interventions, Twenty-first Century Science, which was introduced to support the revised National Curriculum in 2006, illustrates some of the key influences on interviewees' judgements. Following the debate which surrounded the publication of the *Beyond 2000* report, the Twenty-first Century Science approach to the curriculum was introduced embodying the idea of 'how science works'. Although there was a great deal of support for the intervention and its resource materials, it showed how the perceived success of an intervention can be influenced by a variety of factors.

Individual philosophies**:** As one interviewee explained, "Schools that loved it thought it was great but others less so". [I06]. For some teachers the whole approach was considered inappropriate and they resisted its adoption, arguing that it was not doing "proper science" and did not provide a sound basis for study at A-level; as such it was considered a failure. On the other hand other teachers thought it provided a much better approach to science education for all pupils and they argued that, where necessary, they could provide the bridge between GCSE and A-level.

Lack of understanding of the principles underpinning the intervention**:** The introduction of a new approach depends on a clear understanding of the underlying principles and its overall purpose. Failure to get these across can result in the unsuccessful implementation. It was felt that Twenty-first Century Science did not work for those who failed to understand its rationale.

Curtailed development phase**:** The timescale for developing a successful intervention varies depending on its scale and scope. Failing to complete the process or foreshortening it can have negative repercussions. Again, Twenty-first Century Science illustrates this point: several interviewees commented that, although it was felt to be based on some sound thinking with a strong rationale, the results of the pilot were not available to inform its implementation in schools or its roll-out across the country. The effect of curtailed development was also illustrated by the fact that, unlike the curriculum content and pedagogical approach of the programme, the development of appropriate forms of assessment was rushed and restricted.

Impact of views beyond education community**:** Few curriculum interventions get the level of exposure in the press and media that Twenty-first Century Science encountered but it is a vivid illustration of the impact such publicity can have. Although openness of debate is important, the view was expressed that some notable individuals criticised the intervention without having much knowledge of the programme and its aims. Despite gaining some support in the media, the "bad press reduced its success effect" [I18].

Changes in policy environment**:** The direct involvement of governments in what is taught and what happens in the classroom has steadily grown since the introduction of the National Curriculum in 1989. Although it is still used by many schools, Twenty-first Century Science is perceived to have been affected by changes in policy, notably the introduction of the Triple Science programme and, more recently, the drive towards more 'traditional' approaches that are reflected in the current (2013) reforms of the curriculum and qualifications.

Enhancement and enrichment interventions

Enhancement and enrichment (E&E) activities offer opportunities to young people that broaden their experiences of science and help contextualise the curriculum. As curriculum interventions have declined, E&E interventions have apparently increased; this increase reflects the surge in interest in

STEM on economic and workforce grounds, and also, in part, the perceived restrictions in mainstream education due to the National Curriculum and accountability regimes. Thus for many interested parties, the way to ignite young people's enthusiasm about science was through E&E. More and more organisations have become involved in an enormous variety of E&E interventions, ranging from competitions to ambassador schemes and from role models to promotional campaigns. The scope and scale of these interventions varies greatly, from local to national. One of the results of such diversity is that most of the interventions are known to a relatively small number of people, so making judgements as to their success is not at all straightforward.

It was clear during the interviews that unless an individual had had an involvement with an E&E intervention, their perceptions were very much based on anecdotal evidence and hearsay. However, as with the curriculum interventions, the examples of E&E interventions were presented in a constructive manner, with successful elements pointed out before shortcomings. For example, some of the national schemes were recognised as successful in terms of the underlying idea and that they worked very well in some places but were considered ineffectual in others; this variability was seen as a significant limit to the overall level of success. Other areas of doubt cast on large national schemes were the value for money and the lack of sustainability beyond the initial promotion of the activity and resources.

Many local E&E interventions were perceived as successful in their own terms but the criteria for success often seemed limited to whether the young people appeared to enjoy it. It was considered that the majority of the E&E activities were of value but the extent of their value was very difficult to determine – especially, as noted in the NAO report (2010), as many of them are not evaluated at all.

6. Judging success

In discussing the question ‘What is success?’, Section 4 identified the following seven factors that framed the context in which the interviewees and focus groups judged the overall success of the interventions being considered:

- The extent to which an intervention met its objectives.
- The effect on the audience or beneficiaries.
- The principles on which the intervention is based.
- The wider influence of the intervention.
- The return on investment.
- The type, quality and quantity of evidence available.
- The multi-dimensional nature of the judgement.

All participants were very clear that judging the success of an intervention was not ‘black and white’ and that judgements would be influenced by their timing and scope and by the perspective from which they were made.

The timing and scope of the judgement

No two interventions are alike, nor do they follow the same pattern in terms of their development or impact. Similarly the effect of an intervention changes through its life. Thus applying exactly the same criteria to every intervention regardless of its size, scope and stage of development is unhelpful. Some interventions show great promise in the early stages but the effects at best level off and at worst decline and become detrimental. As one interviewee stated, “We don’t have good measures of success and are very reliant on short-term measures” [I02].

Early ‘success’ can be misleading as some of the interventions, usually those of larger scale and duration, were considered to be successful because of the extended influence they had on science education more widely. One interviewee suggested that success criteria might be staged and applied sequentially to reflect the effectiveness at different phases during the intervention and then after its completion. Although such an approach could prove costly and time consuming, it could help provide a more holistic way of judging success and deciding when an intervention needs to be stopped.

The particular perspective from which the judgement was being made

Without exception the interviewees all pointed out that their responses and observations were greatly influenced by their current roles, experience and personal philosophy of education. Despite this, there was considerable consensus around not only the criteria they might use to judge success but also the elements they felt should be incorporated into an intervention in order to give it every chance of being successful.

The number of interviewees involved in this study was too small to explore whether perceptions of success were associated with their different roles, especially given the overlapping roles and previous experience of many of the participants. However, it is possible to demonstrate a general shift in emphasis across the spectrum of roles being undertaken, which ranged from policy makers through funders, science education researchers, teacher educators, curriculum developers, E&E providers and teachers. This can be illustrated by briefly highlighting some of the features identified in three areas of the spectrum: at one end policy makers, at the other teachers, and the far from homogeneous group in between.

Policy makers

Although not insensitive to other aspects of success, the core criterion for policy makers was the need for data which could be used to demonstrate that a specific line of policy was having the desired effect. "Data is fundamental and needs to be used more effectively in order to focus interventions more specifically" [I17] and because of the scale of such interventions, proxy measures have to be used. This often results in the use of simple numbers to monitor effects in order to demonstrate shifts in performance at the population or system level as opposed to the individual or school level.

Adoption of a policy through uptake of an intervention is seen as a key criterion but, as one interviewee pointed out, "Compliance is not really a measure of success" [IC02]. Other changes need to be demonstrated, using other means including accountability measures of school and pupil performance. Whilst recognising there is no direct causal link between a single intervention and the changes that occur, there is always the need to strive to identify some form of relationship, although the NAO report (2010) clearly demonstrates how difficult it is to isolate and demonstrate the effects of a single intervention.

Particularly at a time of economic difficulty and political change, policy makers come under pressure to justify decisions to terminate interventions and to introduce new initiatives that are in line with the philosophy of the government. Neither of these is easy but they highlight the demand for evidence and data, not always used impartially, on which to justify those decisions.

Teachers

If policy makers are working at the population or system level then teachers are working at the individual or school level of their pupils and colleagues – a contrast which was very marked. The teachers interpreted the term 'intervention' specifically to mean activities or actions undertaken in the classroom or school and intended to bring about learning. These ranged from differentiated worksheets to the use of computer-based personalised learning resources and from teacher questioning techniques to pupil-led discussions. Success was therefore discussed in terms of how well the intervention helped pupils come to understand something better and to explain it at a later date.

As the teacher discussions widened they considered the effects of whole school interventions which included, on the one hand, particular schemes of work or courses such as Applied Science and on the other, the processes that the school had introduced to monitor the pupil progress alongside the support mechanisms used to help pupils catch up when needed. Examples of the latter are not necessarily subject-specific.

Discussions with the teachers were also marked by the enthusiasm they showed for E&E activities and the value they put upon them for generating pupil interest in science and related areas of study. Without doubt, there was a desire to help pupils achieve the best possible results in examinations, but when pushed the majority of the teachers admitted that seeing pupils absorbed in an activity, gaining enjoyment and satisfaction from it because they had achieved something they had not done before, was probably their highest priority in terms of whether something was successful or not.

Between the extremes

The responses from these individuals generally recognised the need to take account of more than one perspective. Whilst understanding the need for policy drivers they also wished to acknowledge the effect that interventions have on teachers and their pupils. To varying degrees their responses revealed, more so than with either of the other groups, their own personal philosophies about education and views of how teaching and learning might be improved. Thus reference to views about

learning and their importance in contributing to the success of an intervention occurred more frequently. As one interviewee summed it up, “[Your] personal sphere of knowledge not only provides the basis for the intervention but also for whether or not it is considered successful… [and] what it can be judged against” [I03].

Criteria of success

Everyone made the point that overall success is determined by a combination of parameters, some of which are easily measured but others almost impossible to quantify. The criteria for success that were expressed during the discussions fall into six categories, as shown in Table 1.

Table 1: Criteria of success for interventions

Category	Examples of specific criteria
A. Levels of take-up	Number of schools involved or adopting the intervention
	Number of teachers using the intervention
	Number of pupils reached by the intervention
B. Improvement in attainment	Changes in examination results
	Number and level of qualifications achieved
	In-school test results
	Increases in competence levels
C. Engagement of pupils	Level of uptake post-16
	Leaving destinations e.g. pupils going into STEM careers
	Participation in out-of-school activities
	Attitude surveys of interest in science
	Teacher observations of pupil behaviour
D. Changes in practice	Quality of interaction between learner and teacher
	Levels of inquiry-based learning
	Ease and consistency in implementing the intervention
	Extent to which changes become embedded in practice
E. Changing terms of the debate	Degree of influence on policy
	Impact on practice
	Adoption of the principles of the intervention by others
	Effect on wider behaviours of schools and teachers
F. Value for money	Cost–benefit analysis of the intervention

A. Levels of take-up

In many respects counting the number of schools and teachers that take up a particular intervention or the number of pupils who take part is a relatively straightforward measure of success for many activities. Such information is also a contributory factor in determining the value-for-money criteria (see p. 27), but taken on its own it provides no evidence of the quality of the intervention or of its impact on the learning of pupils.

For many of the interviewees a rule-of-thumb measure of the success of an intervention was the number of teachers they met who talked about it, which was generally felt to indicate that it was considered to be helpful and was either being used or having influence.

B. Improvements in attainment

A major reason for the introduction of an intervention is to raise pupils' attainment and achievement. As such, these would be essential criteria against which to judge the success of an intervention; "just doing an intervention is not enough" [I01]. Thus test and examination results, both internal and external, can be significant criteria.

However, these data should not be taken in isolation and there is a need for benchmarks against which subsequent scores are compared. Too often interventions are introduced without any effort to establish the status of pupils' learning before the activity and so interpreting measures both during and after becomes tentative to say the least. Even if before and after measurements were made, as many

interviewees pointed out, it can be difficult to claim that all increases (or declines) are the result of the intervention.

Difficulties in allocating causality, however, should not be a reason for not using measures of attainment as possible criteria of success. Although, as one interviewee argued, "Assessment is a narrow view" [I16] to take on success, there is a strong case for not only improving the quality of the data collected but also for joining up data sets more effectively. "We don't have good measures of success… and possibly need to make better use of existing structures e.g. the National Pupil Database alongside feedback data [from schools] that can be built into the database to show the effects of interventions [they have adopted]" [I02].

C. Engagement of pupils

Determination of the level and effectiveness of engagement requires an increasing degree of sophistication in the criteria of success. Shifts in quantitative data would include, for example, the number of pupils continuing to study science beyond the age of 16 and destination figures for pupils going into science or STEM related jobs and careers.

More significantly, success criteria in this group have greater emphasis on qualitative evidence than those in groups A and B. The key factor was the way in which pupils react to activities, which, as noted previously, is something that teachers held in high regard. Observations of pupils and teachers responding to particular interventions, demonstrating enjoyment, enthusiasm, inquisitiveness and a desire to do more, all contribute to the perception that an intervention is successful. However, one of the challenges with these criteria, as with those in groups D and E, is the need to gather evidence in a way which is robust and reliable.

Although enjoyment is a starting point, as one interviewee remarked, "[it] needs to be more than 'Did the kids have a good time?'" [I08]. Thus engagement criteria need to look more closely at the way in which pupils are engaging with the activities and subject matter in order to provide specific evidence of whether they were engaged. For example: Did they raise their own questions? Did they talk about it after the session finished? Did they come back at a later date with additional information or ideas about the topic being studied? Importantly, as with the application of all criteria, making judgements about pupils' engagement cannot be taken as one-off events but must be seen over a period of time in order to establish the impact of the intervention in addressing the question posed by an interviewee: "Are we developing more willing learners?" [IC03].

D. Changes in practice

There was a strong consensus that teachers are key to the success of interventions and, by implication, without their involvement and support, an intervention will fail. Thus teachers' reactions to the interventions were regarded as highly important indicators of how successful interventions were perceived to be. One interviewee went as far as to say that "what teachers talk about is the equivalent of success" [I11].

Claims that there has been a change in teaching practice need to be substantiated with specific evidence as to how it has changed. The situation is complex and criteria in this group are dependent on other factors such as improvements in pupils' learning and experience, school performance measures and other external requirements, and differing views of learning. Thus the identification and application of any criteria in this group has to be done with careful consideration but this does not detract from the importance of looking for changes in practice which are likely to contribute to the success of an intervention.

If a change in practice is an intervention's goal, one criterion of success is the degree to which the changed practice remains 'faithful' to that defined by the intervention and the ease with which it can be adopted and implemented (as well as the degree of flexibility that can be tolerated before the practice becomes less effective). An intervention could be considered successful when it was closely monitored and controlled but the effectiveness was reduced when modifications crept in as to the way in which it was implemented or interpreted.

E. Changing the terms of the debate

In a few examples, it was argued that interventions were successful not only when measured by some of the criteria covered in groups A to D but also because they "changed the terms of the debate" [IC01]. In reflecting on the success of interventions thought to have achieved this, the key criterion was the degree to which they were perceived to have influenced subsequent developments, including government policy decisions. Although examples of this seem to be rare, the increasing involvement of government in curriculum development and school accountability structures would imply that an intervention needs to be at least accepted by government to bring about widespread change. The *Beyond 2000* report (Osborne and Millar, 1998), for example, was influential in the revisions of the science National Curriculum in 2006-07.

F. Value for money

The cost and cost–benefit of any intervention cannot be ignored. Major interventions may require substantial investment in the early stages with the benefits accruing over time. Other, in general smaller, interventions show positive returns early in their lifespan for a relatively small cost.

A crucial element in the life of an intervention is the point at which it needs to be terminated or modified in order to maintain its effectiveness. Here a cost–benefit approach is helpful, not just in terms of monetary value but also in terms of time and personnel. Too often there is a sense that an intervention can be made a success by 'throwing more money at it'. This is not the case and as highlighted by some of the interviewees there are examples of projects which had a feel-good factor but a more analytical view suggested they were not good value for money, especially when resources provided to schools free of charge remained in cupboards, unused or even unopened.

Applying the criteria

Even with a full set of clearly defined criteria, judging success is not straightforward; the following issues arose out of the discussions.

- The context for which an intervention is designed and implemented is as important as the overall purpose of the intervention itself.
- There may be a risk that the criteria used for making judgements are those which make the intervention look successful. Therefore there should be a mechanism which helps to benchmark the intervention against other interventions addressing the same problems.
- The timing of the judgement can alter the way in which an intervention will be perceived, depending on the short-, medium- and long-term effects. For example the initial findings of the CASE project indicated only small effects but two years later there were significant outcomes.
- Changes outside the control of the intervention, e.g. in the policy climate, may also affect the way in which an intervention might be perceived.

- The availability of evidence, and the ease with which it can be collected and presented to show the effects of the intervention, can be a significant factor, especially when much of that material is qualitative rather than quantitative.
- The degree to which one criterion is emphasised over others can alter the perception of an intervention significantly. Similarly, the weighting and combination of criteria used to make a judgement can severely affect the extent to which an intervention might be considered successful. In discussing examples of interventions, the phrase "on the one hand… but on the other hand…", or something very similar, was frequently used. For example, the impact of the National Strategies in one discussion was articulated as having helped some individual teachers to improve their skills and therefore was considered reasonably successful in that context. On the other hand, as a body of professionals teachers were made to "jump through hoops" [IC03] by the National Strategies and so these were considered unsuccessful in that context by this interviewee.

Evaluation of interventions

The evaluation of interventions was not raised explicitly but the majority of interviewees referred to it as a way of determining how successful an intervention was. There was complete agreement as to the need for evaluations but major concerns were expressed about their limitations, as follows:

- The selection of criteria used, in the opinion of several interviewees, does not truly reflect the objectives of the intervention and focuses more on outputs rather than outcomes. As a result evaluations can fall into the trap of gathering evidence that is easy to collect.
- The purpose of and the audience for evaluations is not always clear; they can become a routine task, completed to meet the requirements of the funder, rather than to genuinely determine the level of success of the intervention.
- Evaluations tend to be over-positive.
- Evaluations are often done immediately after the end of an intervention, in the worst cases as an 'afterthought', so that many opportunities for evidence gathering are missed and any longer-term effects or evidence of sustainability are never recorded.
- The lessons learned from an evaluation are rarely built on. In part this is because so many interventions end up being one-off projects and so there is little transfer of learning from one to the development of another.

It has been argued elsewhere (Science and Innovation Observatory, 2011) and supported in this study, that a more systematic approach to evaluations is needed.

Use of research evidence

In all the discussions it was felt that there is a dearth of research underpinning claims of what makes interventions successful. The CASE project (see p. 13) was the most frequent example mentioned of a successful intervention based on research and demonstrating clear improvements in pupil performance. The SPACE (Science Processes and Concept Exploration) project[10], which explored primary-aged children's ideas of scientific concepts, was also considered to be a successful research-based intervention. While there was widespread agreement that interventions should have a research base, the over-riding perception was that the vast majority did not. At best, "Most stuff [interventions] is identified as 'good practice' and shared, which is not strictly research" [I10].

[10] Further details are available at: nuffieldfoundation.org/primary-science-and-space (accessed 3 April 2014).

In addition, use of the term 'research' is very loose, with a variety of sources referred to when giving examples including: academic research papers and meta-analyses of studies in this field, evaluations of individual interventions, Ofsted reports, and independent surveys. Although each of these sources of evidence has a role to play and can claim to be 'research', it was noteworthy that none of the interviewees differentiated between them as sources of evidence.

Despite the emphasis placed on the need for robust research and good evidence to develop successful interventions, the perception is that in practice little use is made of either existing evidence or of that which is gathered during the intervention itself. Specific concerns were expressed about teachers' lack of engagement with research. One interviewee was of the view that "use of anecdotal evidence was strong in education", and asked: "Are science teachers scientists? Do they think like scientists?" [I10]. Although many teachers involve themselves in some form of action research, many resist using the evidence available when they feel it does not fit with their own experience. This resonates with previous findings that "even if the evidence was extensive and suggested that adopting a particular teaching strategy could bring about improvement in learning, many [teachers] would not necessarily adopt change. The research evidence had to resonate with their existing perceptions of effective practice. If the teaching strategy was not close to their existing views, they were likely to find reasons for rejecting the research findings" (Ratcliffe, 2010).

More recently a similar finding was expressed in the evaluation of the Pupil Premium: "Schools tended to structure their provision around what their internal evidence told them was needed and what would be effective in tackling disadvantage. This meant that they treated external guidance and research evidence as more or less useful advice rather than as authoritative imperatives. This led many schools to experience some tension between what they believed they were expected to do by external authorities, and what they understood to be in the best interests of their pupils" (Carpenter, 2013, p99).

A frequently mentioned barrier to using evidence more widely and effectively was a lack of opportunity to access it. There is "scant time to look at research. Another thing that hinders progress in education is that researchers get excited but communicate to other researchers so not getting out to a broader field. CLIS was a strong model of what is possible" [I11]. Despite efforts by researchers, this remains a strong, stereotypical perception which emphasises the divide between research and practice. Although there are the tools to access research findings more easily now than previously, the plethora of material is overwhelming. Thus, "the challenge of bridging the gap between research and what happens in the classroom remains" [I02]. This is exacerbated by the fact that "people move on so [there is] no drive to push publications so impact is reduced and lessons [are] not learnt" [I16].

On a more positive note there was acknowledgement that the increased focus in recent years on evidence-based practice and policy was a step in the right direction but there are still many questions to be addressed – for example: "How are interventions constructed? How do theories, including those from beyond education, inform interventions?" [I07] – as well as logistical and philosophical concerns to overcome such as the use of randomised controlled trials, the matching of large and small-scale practice, and greater understanding of the processes required for developing successful and effective interventions.

7. Developing successful interventions

Throughout the interviews and focus groups there was a unanimous view that there is no silver bullet which will ensure the success of an intervention. There was, however, a consensus that successful interventions require a combination of elements underpinned by individuals with the appropriate combination of experience, knowledge and skills. Finally, there is a degree of chance in that some interventions appear to "catch the tide" and the "mood of the time" in relation to the political environment or the appetite for change within the education community. This section considers these matters in some detail before examining perceived barriers to success.

Factors that contribute to success of interventions

The factors identified as contributing to successful interventions echo the different types of criteria discussed in Sections 4 and 6. However, there is an important distinction between the two. Criteria include both qualitative and quantitative measures or indicators of success, usually in the form of outputs and outcomes that provide evidence of what has, or has not, been achieved. The factors which contribute to success are the inputs to a process which has been (or should have been) designed to meet the objectives that have been determined for the intervention. The discussion which follows therefore reviews the factors from the perspective of how they contribute to the effectiveness of the process of the intervention. In this context successful interventions were perceived to depend on the following six elements.

Definition of the purpose of and need for the intervention

Interventions should have a "clear and well developed rationale – what is distinctive and what will it contribute; good ideas are not enough" [I12]. There needs to be a "clear purpose linking to the activity" [I14] and "intervention is needed to solve a problem" [I16]. "Kids need to see the point" [TF1]. These comments reflect the consensus that a successful intervention should have a clear purpose and meet a recognised need.

The terms 'purpose' and 'need' were used almost interchangeably by most interviewees, yet there is a distinction that is often missed, contributing to the failure of some interventions. In general terms an intervention that has a well-thought-out purpose but takes little or no account of the needs of the target audience is more likely to fail than if both are aligned. For example, Stimulating Physics, which was seen to be successful by many interviewees, had a clear purpose of increasing the number of young people taking up physics but it also met a need of teachers by improving their knowledge of physics and confidence to teach it.

Both purpose and need can come from a variety of stakeholders and depend on their particular areas of interest. For a researcher it may be important to explore a particular idea in more depth, providing both a theoretical perspective and an evidence base for the intervention. For a teacher, the purpose and need are likely to have a more practical focus that will help provide a solution to a specific problem. Some teachers felt frustration about interventions which, from their perspective, failed to meet their needs or those of their students: "Initiatives are too often top down rather than built up from the bottom" [TF2].

Attention also needs to be given to addressing the questions "Why are we moving in the direction determined by the intervention?" [I05] and "Are the questions being asked the 'right' ones?" [I05]. It

would appear that the purpose of and need for interventions are not always fully thought through; as one interviewee (perhaps unkindly) suggested, there is the "civil service aspect of education – the tendency to grab a good idea and run with it" [I11] when there is a need for more evidence and greater analysis of the situation and possible consequences.

Clarity of the process

For many of the interviewees the clarity of the process for initiating, designing, implementing and reviewing an intervention was considered to be more important to its success than having a well-defined purpose and need.

Ideally an intervention would have an overall plan which was made explicit to all those involved clearly setting out, the 'what', 'why', 'where', 'when' and 'how' for each phase of the project. Initially this might be at quite a high level. As the intervention progresses, each phase becomes more defined and refined, adapting to changing circumstances, the rate of progress and the impact of the intervention itself as it proceeds, whilst staying true to its original purpose. "Flexibility needs to be there for successful interventions" [I17].

Collaboration, effective dialogue and clear lines of communication are all important. As one interviewee complained, "Not enough consultation goes into the development of initiatives" and "you shouldn't do interventions to people" [I16]. Others made similar comments about the need to involve teachers in the processes of setting up and conducting interventions.

Effectiveness of the implementation

Closely linked to the clarity of process is the importance of the quality and effectiveness with which an intervention is implemented. It is particularly necessary to "work with what is out there – existing infrastructure" [I19] because only very rarely, if ever, can an intervention change the existing landscape or isolate itself from the day-to-day pressures, attitudes and behaviours. Thus implementation plans must not only address the practical tasks and logistics of project management but also tackle the challenges of changing practices. As one interviewee explained, "it must not be simply a set of procedures… success is where teachers have taken time to understand what [the intervention] is about" [I03]. Particular attention needs to be given to different phases of the life-cycle of the intervention. These include the following:

Initiation: An intervention needs to be introduced at the "Right time and right place so it addresses the problem when it is needed" [I06]. This should take account of other things that are happening at the time, which may result in distractions or other activities being prioritised over the intervention in question.

Development: Getting the correct balance between maintaining the momentum of an intervention and the need to ensure that all parties are kept on board is not easy. Making progress is important but things need to be consolidated too, as one of the teachers reflected: "The National Strategies were delivered mechanistically and too fast" [TF2] – short-term issues often overshadowed the longer-term goals.

Interventions need to be tested and "operationalised so they can be taken up in the classroom" [I12] or, put another way by the same interviewee, "we need to put the intervention in harm's way". The development phase is when it must be tested in the environments where it needs to succeed rather than being a laboratory exercise in which variables have been controlled to an extent that the intervention is cushioned from competing priorities and developments.

Scaling up: Scaling up an intervention is seen to be a major challenge, whether encouraging a colleague to take it up, embedding it across a school or getting it adopted across a regional or national system. In part this depends on other people understanding the need for the intervention and recognising that it can address the problems they have. At the core there is a need to get people talking about it and using it in order to build up a critical mass of users so that it becomes part of ongoing practice.
Issues of promotion and dissemination, as discussed below, are important but other factors also need to be considered. Perhaps the most important is to ensure that teachers are supported to understand the basis for the intervention and its implementation. Just telling them how to do it is not enough. Opportunities – structure and time – are needed to share experiences and feedback.

Some interviewees noted that it would be inappropriate to try to scale up an intervention beyond its natural reach. As one interviewee put it, "You can scale up McDonald's but not five-star Michelin restaurants" [IC02].

Sustainability: Many projects appear to be successful in the short term but have little life beyond their funding period. For some this is planned, but few seem to have an extended life beyond the original funding phase. Of the examples referred to in this study, the CASE project was the only one considered to have such longevity while retaining, to a large extent, its original identity. Others, notably the Nuffield curriculum projects and the SPACE project, were considered to have influenced subsequent thinking and developments, thus leaving a legacy rather than being sustained as discrete interventions.
It was generally agreed that not all interventions are or should be sustainable in their own right but too often the question of sustainability is not considered at the outset or built into the design of the intervention. One interviewee suggested there was a need to consider building a virtuous circle of success using feedback to provide the foundation for the next step or refreshment of an intervention.

Considerations about sustainability need to take account of the wide range and scale of interventions covered. In the majority of cases, it is assumed that sustainability means the same intervention extended over a longer period, but sustainability may take other forms: a legacy of influence, a new way of providing wider experiences for young people, or greater confidence and increased expertise for teachers.

Applicability to new situations: Most interventions are designed, developed and tried out in a particular context and therefore could be considered successful in some settings but not in others. Depending on the objectives of the intervention, some consideration should be given to the effectiveness and value of the intervention in other settings and to different groups of young people. Issues such as gender, cultural background and socio-economic status are all perceived to influence the success of an intervention, as can the challenge of engaging hard-to-reach pupils.

The people involved

"People make them work" [IC04], "People give impact" [IC02] and "People are central" [I19] are three statements which reflect the unanimous perception that the success of any intervention depends ultimately on the quality of the people involved. Interventions involving good people can fail but it is much less likely than an intervention being successful without good people. With few exceptions, successful interventions were perceived to involve a team of individuals fulfilling complementary roles with different skills, including technical skills (in, for example, research or project management), communication skills and, importantly, leadership skills.

Having a strong team is only part of the challenge because, as was frequently pointed out, there are other groups of people who need to be involved. "Getting teacher involvement is absolutely vital" [I07] for a range of reasons, including to "get a sense of how [the intervention] will be received by teaching community" [I13], to "build time for teachers' professional learning… [and] help them make sense of the intervention – not just following a recipe or set of instructions" [I04], and to "involve teachers looking for evidence which is too often done by researchers so teachers end up unprepared to continue the approach" [I04].

One interviewee cautioned: "[You need to] look for practitioners – excellent teachers (for example National Teaching Fellows[11] or members of Primary Science Teaching Trust, Primary Teachers College[12]) – if good teachers can't or don't do it, what chance is there for getting weak teachers involved" [I16]. The corollary of this is that interventions have to be implemented by teachers with a wide range of expertise, so an intervention might work with highly skilled teachers but be less effective with less experienced ones – hence the need for specific professional development to be built into the intervention.

Senior leaders in schools must take the intervention seriously and other stakeholders need to be involved at an appropriate level in order to maximise its chances of success. An important group of stakeholders is the pupils themselves who, when given a voice, can provide insightful observations and vital feedback. They too need to see the benefits of an intervention.

Evidence, monitoring and accountability

As discussed in Section 6, interviewees all expressed a need for evidence on the success of interventions, yet they also expressed concern that in general there was little robust evidence on which to build. One observed: "Too many interventions are done because they are a 'good thing'" [I07]. It was felt they are more likely to be successful if there is an evidence base on which to initiate the intervention, to monitor its progress and impact, and to account for the investment put into it.

The evidence and measures used for monitoring and accountability need to relate specifically to the overall objectives of the intervention as well as the context in which it is implemented. As one interviewee argued, "Successful interventions need to speak to teachers' personal characteristics, places in which they (pupils and teachers) work, external context including policy accountability, and impact on teaching and learning" [I13]. As this implies, embedding monitoring, evaluation and accountability differs in detail from situation to situation but the involvement of senior leaders and teachers is crucial. Practitioners should be involved in noting changes, collecting evidence, interpreting the data and reflecting on the impact this has on their own practice and, most importantly, on the engagement, learning and attainment of pupils. Overall the perception of the interviewees and teachers was that the quality of the evidence base and the processes for monitoring and evaluation are weak links in the chain of developing successful interventions.

Communication, promotion and profile

An intervention cannot be perceived to be successful if no one knows about it. Conversely, just because a large number of people know about an intervention does not mean that it is good. Thus

[11] National Teaching Fellowship Scheme. More details at: www.heacademy.ac.uk/ntfs (accessed 3 April 2014).

[12] PSTT (formerly Astra Zeneca Science Teaching Trust), Primary Science Teacher College. More details at: pstt.org.uk/science-teaching/primary-science-teacher-college.aspx (accessed 3 April 2014).

developing a project which is seen to be successful requires a communication plan to inform its potential audience, which might include departmental colleagues, all teachers, parents, pupils, policy makers and other stakeholders. With increasing competition for attention in the education marketplace, getting the communication plan right plays a more and more significant role.

The communication plan should be incorporated at the planning stage because it will be an important lever for getting the intervention embedded into ongoing practice, building a strong profile and establishing its credibility. Thus there needs to be:

- clarity of message, based on evidence, with which the target audience identifies because, as some interviewees pointed out, evidence on its own is not enough
- a variety of mechanisms and channels of communication
- enthusiasts for the intervention who will champion it but not oversell it.

The importance of word-of-mouth dissemination cannot be overestimated, nor can the influence of social media in its different forms be ignored. Teachers talking to teachers, whatever the medium, is still seen to be a critical mechanism by which an intervention is adopted and so perceived to be successful.

Barriers and challenges to success of interventions

If the above factors are not adequately in place then an intervention's likelihood of success is much reduced. There are other factors which in the right circumstances can be considered to contribute to the success of an intervention but more often are perceived to be barriers. The majority of those mentioned during the interviews and focus groups relate to three broad concerns.

System issues at national level

All participants expressed major concerns about the degree to which national policies can severely impact on the success of an intervention, including the large number of changes brought in over recent years giving little time to review the effect of an intervention. Furthermore, the sheer number of interventions that exist makes it impossible to identify which has made a particular impact.

Accountability and the way in which it influences the criteria used to define the success of an intervention was most frequently identified as problematic. "Accountability measures influence the impact of interventions" [I20] and "Interventions are devised to meet the rules of the game; playing to whims reduces the chances of success" [I15] reflect the perception of this increasingly problematic issue. The fixation on targets and accountability measures is perceived to have altered behaviour of not only teachers but other stakeholders and funders too, so that success of an intervention has become determined by how well it helps to meet a specific target or measure, curtailing the development and trialling of interventions which do not show immediate improvements in pupil performance, as measured by the accountability measures. Such practice does not necessarily bring about a fundamental, long-lasting improvement in teaching and learning. Although one interviewee observed: "Performance tables don't have the power people think they have – attitudes and beliefs are much more critical" [I15].

System issues at school level

Even in schools where there is a strong desire to be innovative and to explore different types of intervention, the continual tension between the external demands, the internal requirements and the desire to 'push the boundaries' is perceived to be restrictive. "There is a need to be aware of the

bigger picture and get a balance between exam results and interest and engagement but the balance gets out of kilter. The short-term issues often overshadow the longer-term goals" [TF2]. "The main barrier is the climate in schools – there is too much out there, senior managers need to be convinced. Too much seems to be prescribed so teachers don't have time to think outside the day-to-day. This restricts opportunities and reduces potential" [I02]. As such, "Many schools are not learning organisations and don't prioritise the learning of teachers. There is top-down drive of compliance of external mandates" [I12]. This type of climate reduces the likelihood of interventions being successful beyond the targets and accountability measures.

Further tensions, this time between whole-school matters and subject issues, were also perceived to hamper the development of successful interventions. This was felt to be particularly true of primary schools, especially when science was not a priority in the school development plan so there was no budget for development. In secondary schools this was felt to be less of a barrier, but nevertheless, if the departmental culture clashed with that of the school, the likelihood of success was reduced.

Positions that people adopt

The ways in which different individuals perceive an intervention and relate to it are critical in contributing to its success. One interviewee argued: "All barriers are in one's head, so there is a need to get hearts and minds engaged in order to overcome resistance to change" [I19].

Although it was acknowledged that the commitment of the senior management team could be a big driver towards success, the most frequently expressed concern was expressions of frustration that senior managers too often block interventions. In primary schools, mainly because of their size, this was seen as more of a personal matter, with senior management teams being so heavily involved in the day-to-day teaching whereas in secondary schools senior management teams were felt to be more remote.

8. Towards a framework for successful interventions

As argued in Section 4, there is no black and white distinction between what is perceived to be a successful intervention and what is not. Ultimately it is a judgement based on a combination of factors. This study has attempted to gain some insights as to how a range of individuals perceive interventions to be successful or not. In doing so it has highlighted their views from two angles: the criteria they use to judge success (see Section 6) and the factors they consider to contribute to developing a successful intervention (see Section 7). Not surprisingly, how the criteria would be applied and the degree of emphasis given to each of the factors will depend in broad terms on the nature and scale of the intervention, its objectives, the context in which it is conducted and the personnel involved. Despite the wide range of views in detail, there is a significant degree of consensus that is reflected in this report and from which arise seven cross-cutting issues giving grounds for further work in this area. This section highlights these issues, with recommendations, and endeavours to set out the basis of a model for the development of successful interventions.

The clarity of purpose and shared understanding of a successful intervention

Successful interventions are perceived to have a clear purpose and objectives, both short- and long-term, that are shared with and understood by all stakeholders. At the same time it is felt that many interventions are based on a good idea but their objectives are not well thought through and are rarely challenged in order to determine whether or not this is the right thing to do and the right way to do it.

The majority of interviewees repeatedly stated that successful interventions must meet defined needs, especially those of teachers and of pupils. But focusing exclusively on a specific need can lead to a blinkered view of the problem, divorced from other developments and, potentially, from the real underlying cause – what might be referred to as 'sticking plaster syndrome'. It is interesting to note that many interventions which were considered successful have strong theoretical foundations that challenged existing paradigms. Emphasis was placed on conveying the principles which were then translated into specific activities for the classroom and other learning environments.

The need to develop a wider and deeper shared understanding of what an intervention is trying to achieve and why was emphasised throughout. Clearly this is critical for all those who are directly involved but, as was frequently pointed out, this shared understanding must involve other stakeholders. Unfortunately, the overall perception was that such understanding was often missing, resulting in mixed messages and superficial implementation of interventions.

The differences between reactions from the teachers in the focus groups and those from other interviewees was very marked. It was clear that a communication gap remains between teachers and the originators of interventions on all scales.

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.*

The quality, quantity and nature of evidence used to define and judge success

One of the reasons for this study was a concern about the lack of evidence that exists in relation to judging the success of interventions. The findings add to the much wider debate about the use of evidence to inform teaching and learning practices and the mechanisms by which they are introduced and adopted. As discussed in Section 6, the evidence base for the effectiveness of individual interventions is not, in general, very strong, despite the fact that there is almost universal acceptance that robust and valid evidence is critical in deciding the success of an intervention.

Strengthening the evidence base should be a high priority in addressing the effectiveness and success of interventions. However, as with other terms, the word 'evidence' is used to cover a multitude of things and its meaning is not always clear. In the context of this study, three particular issues arise.

The first is the lack of distinction being made between the different reasons for collection and use of evidence. Just as there needs to be greater clarity in the purpose of the intervention, there also needs to be greater clarity in the way evidence is collected and used to assess the effectiveness of an intervention. The range of evidence required goes from 'pure research' data testing to information on the way in which funding has been spent. All forms of evidence have a value and place but only if they match the question being asked. In many of the conversations there was no distinction made about the types of evidence that are needed to address the different questions that are being asked.

The second issue concerns the lack of well-framed questions to which answers are being sought. Too often, evidence that is easily measured or counted is collected without having defined the question to which the evidence might help provide an answer, especially with low quality evaluations.

The third issue is the need to be able to recognise unexpected outcomes, both positive and negative, which may result from the intervention.

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.*

The degree to which the situational context affects the likely success of an intervention

There is a strong view that the situational context – for instance, the backgrounds of the pupils or the location of the school – into which an intervention is introduced influences its likelihood of success. When schools are part of a project the intervention is more likely to succeed. This perception raises the questions of the extent to which interventions have to be tailored to specific contexts or groups of pupils and, conversely, the extent to which an intervention that has been shown to be effective in some contexts can be modified before it ceases to be effective.

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.*

The extent to which the impact of policy changes might hamper or support the initiation and development of successful interventions

The policy environment is considered to have had an increasingly strong influence on the number and nature of interventions, either through direct funding of specific activities or the encouragement of particular programmes of work. Science, and more generally STEM, has been the focus of strong central influence in the last 10 to 15 years, because of its link with the economy, and is probably the only curriculum area in England that in the current economic climate has any central funding. Science (and STEM) education also has probably the highest number of stakeholders (ranging from learned societies and professional bodies to large companies) trying to add their own interventions into the mix.

This raises two main areas of concern. The first is exemplified by the question: to what extent are interventions designed and developed to either 'chase the funding' or 'meet accountability targets' rather than to support high quality teaching and learning? Clearly the different objectives are not incompatible but the policy environment does affect the balance. Taking a longer-term view, it would appear that the opportunities for substantial curriculum development interventions have declined since the introduction of the National Curriculum in 1989.

The second concern relates to the plethora of interventions that exist. From a school's perspective, the sheer number makes it difficult to plot a way through and select the most appropriate interventions. The problem is compounded by difficulty in isolating the effects of a particular intervention on the engagement and achievement of pupils or on the effectiveness of the teaching and learning. The interactions between different types of intervention are very difficult to identify but in too many circumstances the potential interference of one intervention with another is not considered.

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.*

The challenge of implementing interventions successfully

As argued in Section 7, the way in which an intervention is implemented, in all its phases, is considered central to its success and effectiveness. Importantly, putting interventions into practice "must not be simply a set of procedures… success is where teachers have taken the time to understand what it is about" [I03]. Ensuring that participants understand the spirit and values underpinning the intervention and have the knowledge and expertise to implement it effectively takes time, which is not always allowed for. One of the teacher focus groups was very explicit in its observations on how the implementation of major interventions failed to be effective. Too many of these adopted top-down approaches and provided inputs that were isolated from the bigger picture and the rationale behind the

intervention. Lead-in times for implementation in school were perceived to be far too short and, all too often, there is a lack of ongoing provision (time, training and resources) for continuous improvement in practice.

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, implementing and evaluating 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.*

The extent to which effective change management might contribute to successful interventions

Many of the interviewees made reference to the need to recognise that interventions are integral elements in bringing about changes in practice and so "require a commitment to make that change". [I03]. This in turn involves change agents and leadership at all levels to "make changes and embed them into ongoing practice" [I01].

Although it was not stated explicitly, there was a perception that wider issues, such as the processes of managing change, are rarely considered in planning interventions or how an intervention might contribute to wider changes that are being introduced in a school, at local or at national level. For interventions to be given a chance of being successful, the potential benefits of the new approach need to be appreciated, people need to be brought on board, and there needs to be time to consider objections, anticipate challenges and ensure the highest possible quality of implementation. For many interventions these issues are not fully explored at any stage, to the detriment of the overall success of the intervention.

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.*

The need for a more holistic model for developing interventions

No intervention is perfect. The complex combination of factors influencing the outcomes and the diversity of criteria by which they might be judged render perfection impossible to achieve. However, based on the discussions presented in this report, there is scope to raise the quality of interventions. Some recommendations have been included in the discussion of the cross-cutting themes set out above. In addition, consideration should be given to adopting a more holistic approach to developing interventions.

The majority of interventions, small or large, short- or long-term, come about because a group of people in a particular situation wish to introduce something which they hope will benefit another group of people, directly or indirectly. Many of these interventions can in fact claim to be successful but too often the success is limited, misrepresented or poor value for money. There are many reasons a particular intervention may be considered to have been less than successful, but the views expressed in the current study suggest that in part it is because too many interventions fail to recognise the complexity of the endeavour. What follows is an attempt to bring together the lessons derived from

the discussions and to form the basis of a model for developing successful interventions by setting the different elements into an overall context.

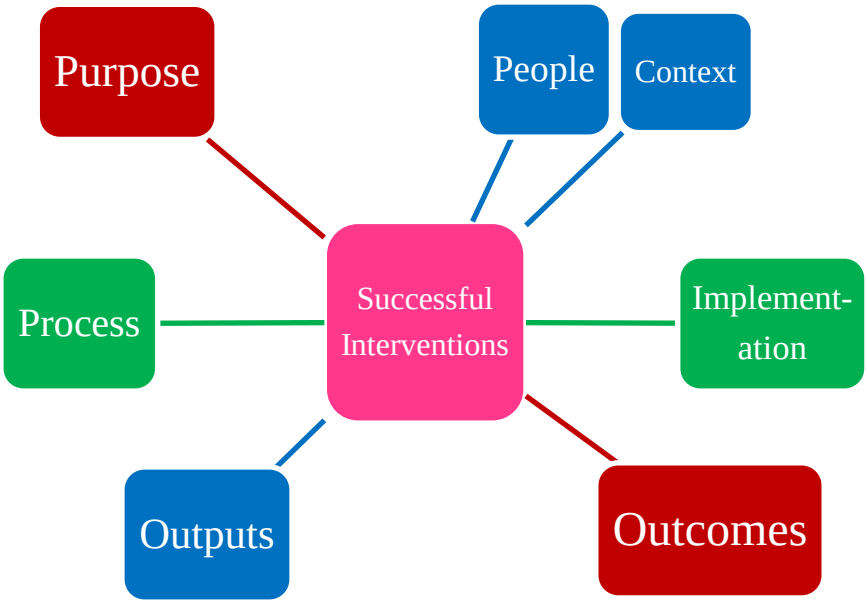
The proposed model envisages seven key elements – purpose, people, context, processes, implementation, outputs and outcomes – that contribute to a successful intervention, as summarised in Table 2 below.

Table 2: Elements of successful interventions

Successful interventions have a clear purpose which:	• matches a defined need (or needs)
	• is underpinned by shared values and principles
	• has an evidence-based rationale
	• is clearly communicated at all stages.
Successful interventions are undertaken by people who:	• have skills in leadership and management
	• have expertise and subject knowledge in pedagogy, curriculum and assessment
	• work collaboratively as a team and in partnership with other stakeholders
	• engage in the necessary continuing professional development to understand and implement the intervention effectively
Successful interventions take into account the context by:	• using evidence to identify the starting points and influential factors
	• building on the existing strengths
	• mitigating weaknesses
	• adapting to local factors
	• engaging teachers in the design and implementation.
Successful interventions establish processes which:	• enable robust and constructive dialogue
	• facilitate clarity of communication
	• allow for effective management of change
	• use evidence rigorously at all stages
	• build a respected profile for the intervention
	• ensure that high standards are maintained across all aspects of the intervention
	• take account of unexpected outcomes.
Successful interventions depend on effective implementation which:	• is well planned with appropriate milestones
	• allows for changes in circumstances
	• monitors, and systematically provides feedback on progress, failures and achievements
	• incorporates regular review cycles and acts on feedback information.
Successful interventions result in outputs which:	• are based on criteria specifically related to the objectives of the intervention
	• are definable and measurable
	• include short-, medium- and longer-term criteria appropriate to the stage, scale and context of the intervention.
Successful interventions bring about change through their outcomes which:	• provide evidence to demonstrate sustainable impact on engagement, teaching and learning

	• add to the evidence base and understanding
	• improve existing practice
	• inform practice in new contexts
	• are fully evaluated and provide feedback for future interventions.

Figure 1: A model for successful interventions



The seven elements all interact and contribute to the success of an intervention but, together, they can be considered to form three intersecting axes, as set out in Figure 1, in which:

- the clarity of, and commitment to, the **purpose** lead to tangible impact and **outcomes**
- suitable **people** working in the right **context** results in measureable and demonstrable **outputs**
- robust **processes** lead to effective **implementation**.

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.*

Although perceptions of what makes interventions successful vary, the findings of this study suggest there is a broad consensus on the key elements required and the issues that are outstanding. Further work is required, however, to refine a framework for developing more successful interventions and to establish robust and reliable evidence to support claims of success.

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.

Adey, P. S., Shayer, M., & Yates, C. (1989). *Thinking Science: The Materials of the CASE Project*. London: Nelson.

Bennett, J., Hogarth, S., & Lubben, F. (2003). A systematic review of the effects of context-based and Science-Technology-Society (STS) approaches in the teaching of secondary science. Research Evidence in Education Library. London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.

Bennett, J., Lubben, F., Hogarth, S., & Campbell, B. (2004). A systematic review of the use of small-group discussions in science teaching with students aged 11-18, and their effects on students' understanding in science or attitude to science. Research Evidence in Education Library. London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.

Black, P., & Wiliam, D. (1998a). Assessment and classroom learning. *Assessment in Education: Principles Policy and Practice, 5*(1), 67.

Black, P., & Wiliam, D. (1998b). *Inside the black box: Raising standards through classroom assessment.* London: GL Assessment.

Carpenter, H., Papps, I., Bragg, J., Dyson, A. Harris, D., Kerr, K., Todd, L., & Laing, K. (2013). Evaluation of Pupil Premium: Research Report. London.

CUREE. (2011). Implementing curriculum change - A synthesis of the evidence from curriculum change 2007-2010. Coventry,UK: Centre for the Use of Research and Evidence in Education.

Department for Education and Skills, Department of Trade and Industry. (2006). The Science, Technology, Engineering and Mathematics (STEM) Programme Report. nationalstemcentre.org.uk/res/documents/page/stem_programme_report_2006.pdf (accessed 3 April 2014)

Harrison, M. (2012). *Jobs and growth: the importance of engineering to the UK economy*. London: Royal Academy of Engineering. raeng.org.uk/news/publications/list/reports/Jobs_and_Growth.pdf (accessed 3 April 2014).

Higgins, S., Katsipataki, M., Kokotsaki, D., Coleman, R., Major, L.E., & Coe, R. (2013). The Sutton Trust-Education Endowment Foundation Teaching and Learning Toolkit. London.

Higgins, S., Kokotsaki, D., & Coe, R. (2011). Toolkit of strategies to improve learning. Summary for schools spending the pupil premium. London: The Sutton Trust.

Hogarth, S., Bennett, J., Lubben, F., Campbell, B., & Robinson, A. (2006). ICT in science teaching. Technical report. Research Evidence in Education Library. London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.

Johnston, R. S. W., Joyce, E. (2005). A Seven Year Study of the Effects of Synthetic Phonics Teaching on Reading and Spelling Attainment. Edinburgh.

Jones, M., & Gott, R. (1998). Cognitive acceleration through science education: alternative perspectives. *International Journal of Science Education, 20*(7), 755-768.

Lubben, F., Bennett, J., Hogarth, S., & Robinson, A. (2005). A systematic review of the effects of context based and Science-Technology-Society (STS) approaches in the teaching of secondary science on boys and girls, and on lower ability pupils. Research Evidence in Education Library. London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.

Mourshed, M., Chineze, C., & Barber, M. (2010). How the world's most improved school systems keep getting better. London: McKinsey & Company.

NAO. (2010). Educating the next generation of scientists. London: National Audit Office.

Osborne, J., & Millar, R. (Eds.). (1998). *Beyond 2000: Science education for the future. The report of a seminar series funded by the Nuffield Foundation*. London: King's College London, School of Education.

Ratcliffe, M. (2010). How science teachers use research evidence. [Summary]. *Better: Evidence-based Education, 2*(3 Spring 2010), 2.

Schroeder, C. M., Scott, T. P., Tolson, H., & Huang, T.-Y. (2007). A meta-analysis of national research: Effects of teaching strategies on student achievement in science in the United States. *Journal of Research in Science Teaching, 44*(10), 1436.

Science and Innovation Observatory. (2011). Evaluating STEM Initiatives: Are STEM evaluations making a difference and can we make them work better? *Science and Innovation Observatory Policy and Strategy Briefing Papers*. Sheffield: Sheffield Hallam University.

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

Shayer, M., & Adey, P. S. (1992). 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.

Slavin, R. E., Lake, C., Hanley, P., & Thurston, A. (2012). Effective Programs for Elementary Science: A Best-Evidence Synthesis. The Best Evidence Encyclopedia. bestevidence.org/word/elem_science_Jun_13_2012.pdf (accessed 3 April 2014).

Springate, I., Harland, J., Lord, P., & Straw, S. (2009). Evaluation of the 2008-09 DCSF-funded Specialist Schools and Academies Trust STEM Pathfinder Programme. Slough: NFER.

Appendix 1 Interview briefing notes

BRIEFING NOTES AND PRE-INTERVIEW QUESTIONNAIRE

A statement of the problem

There is a large number of interventions designed to improve teaching and learning in schools but there is little understanding of why some are considered to be successful and others not. Often, it seems there is only sketchy research evidence to support the perceptions yet decisions are being taken at national, local and school level on which to adopt and promote.

Purpose of the study

The purpose of the current study is to explore the perceptions of a range of stakeholders in order to better understand why some interventions are considered successful and so are more likely to be promoted or adopted, to test whether these perceptions are supported by existing research evidence and to identify any common elements which contribute to successful interventions. Ultimately the findings could inform the development of future interventions and the quality of decision-making.

Explaining the use of the principal terms

As the purpose of the study is to explore individuals' understanding and perceptions it is important to avoid being over-restrictive in defining terms. At the same time it is necessary to clarify a common view as to the meaning of the principal terms as used in the study. Thus,

> **Intervention** is used to refer to programmes and activities that aim to improve teaching and learning in order to raise achievement and improve learning experiences in science education. It includes activities designed to alter approaches to teaching and ways in which the curriculum is implemented. For the purposes of this study therefore, the term **intervention** will be used generically to cover a wide range of actions and activities.
>
> **Successful** will be used as an all-embracing term to describe the achievements of an intervention and can include a wide range of criteria. The actual criteria or measures used are likely to vary according to the role and position of the interviewee. Indeed, the way in which individuals describe success in relation to an intervention is part of the study.

Purpose of the interview and pre-interview questionnaire

The purpose of the interview is to explore:

i. what you consider to be / have been successful interventions;
ii. the factors that you consider contribute to making an intervention successful;
iii. the criteria by which you think success might be judged;
iv. the relationship between perceptions of success and existing research evidence;
v. the challenges involved in extending the influence of interventions so they are adopted more widely and made sustainable.

The pre-interview questionnaire which is follows consists of filling in the table. It is simply to save time in the interview by gathering information on interventions you consider to be / have been successful. Those already listed are ones that were identified in a previous round of 'structured conversations'. The list is certainly not exhaustive and you are invited to add as many as you wish.

The interview will use the completed questionnaire as a starting point for the conversation so I would be grateful if you could return it to me before we talk.

PRE-INTERVIEW QUESTIONNAIRE

1. Please complete the table below by placing a tick in the appropriate column for each intervention.
2. Add to the end of the list any interventions you would describe as successful.

Please feel free to add any comments /observations but note these are not required as they can be picked up during the interview.

NB. There are some interventions, listed at the end, that are not science specific but in the previous conversations were felt to have had some influence on science teaching and learning.

Intervention	Don't know it	Consider successful	Consider unsuccessful
National Strategies in England			
Triple Science Programme			
Network of Science Learning Centres			
How science works			
Stimulating Physics			
CASE			
Nuffield Projects (referred to as a group)			
SATIS[13]			
Suffolk Science			
Science Enhancement Programme			
SPACE[14] / Nuffield Primary Science			
Twenty-first Century Science			
Nuffield STEM Project			
Wellcome Trust Darwin projects			
Wellcome / Camden STEM Project			
Reachout Lab Imperial College			
Assessment for Learning (AfL)			
SureStart			
SEAL[15]			
Thinking through Geography			

Comments, if any. Please continue overleaf.

[13] SATIS: Science and Technology in Society, originally developed and published by The Association for Science Education.

[14] SPACE: Science Processes and Concept Exploration Project; Kings' College London and Centre for Research in Primary Science and Technology (CRIPSAT), University of Liverpool.

[15] SEAL: Social and Emotional Aspects of Learning Project (see for example: Department for Children, Schools and Families (2007). *Social and emotional aspects of learning for secondary schools.* Nottingham: DCSF Publications.)

Appendix 2 Interview template

SUCCESSFUL INTERVENTIONS TO IMPROVE TEACHING AND LEARNING IN SCIENCE: exploring perceptions

INTERVIEW TEMPLATE

INTERVIEWEE:	Date:
	Telephone:

SECTION 1a:.
To identify examples of interventions in science education which the interviewee considers to be successful and other interventions, if any, that may not be directly science but have had an influence on science education.

Questions	Prompts	Comments
Responses recorded on pre-interview questionnaire	Remind interviewee of openness of the definitions.	The purpose of this question is to get the interviewees perceptions of successful interventions and so the question and definitions have been deliberatively kept open.

SECTION 1b:
To outline reasons why the interviewee selected the interventions given in 1a

Questions	Prompts	Comments
Why did you consider these interventions (given in response to 1a) to be successful? Can you give specific reasons why you considered X or Y to be successful?	Encourage specific linking of reasons to particular interventions.	The response to this question may be approached generally but it is important to get the interviewee to consider at least one specific intervention from their list.

SECTION 1c:
To indicate any research evidence the interviewee is aware of to support the examples given in 1a.

Questions	Prompts	Comments
Are you aware of any research evidence to support your view that the interventions you stated are successful? To what extent do you think research evidence is important in order to decide whether an intervention is successful or not?	Depending on the response encourage examples of types of research evidence that is known / considered important in informing decisions.	

SECTION 2:
To elicit the criteria the interviewee would use to judge the success or otherwise of an intervention.

Questions	Prompts	Comments
What criteria would you use to decide whether an intervention was successful or not? What kind of evidence would you look for?	Encourage to be specific and to feel free to use whatever type of evidence is felt appropriate.	Probe the use of qualitative and quantitative measures.
SECTION 3: To establish what elements the interviewee considers are needed to ensure that an intervention is successful.		
Questions	Prompts	Comments
If you were designing an intervention what do you think you need to do to ensure it was successful?	Possibly need to consider planning, clarity of objectives, context in which intervention is to be used, ways in which the intervention is promoted etc.	This question may need to be addressed using a particular example.
SECTION 4: To identify what challenges the interviewee considers there to be in implementing and getting interventions adopted more widely and sustainably.		
Questions	Prompts	Comments
What do you think are the barriers to an intervention being successful? How might you overcome them?		This is a big open question so is unlikely to be answered fully but it should be possible to elicit some examples of the types of barrier that are perceived.

Appendix 3 Interventions

The following table provides a full list of the interventions that were referred to during the interviews for the study. The allocations to an intervention group, as discussed in the main text, is not rigid as many of the interventions span more than one group.

Intervention	Knowledge building	Policy driven	Pedagogy	Curriculum	Enhancement & Enrichment
CASE	X				
SEAL	X				
Nuffield STEM Project	X				
Beyond 2000	X				
Children's Learning in science	X				
SPACE / Nuffield Primary Science	X				
APU Reports for teachers	X				
ASE Be safe	X				
Accountability measures		X			
National Strategies in England		X			
Triple Science Programme		X			
Network of Science Learning Centres		X			
How science works		X			
National curriculum		X			
SureStart		X			
PSQM		X			
Stimulating Physics			X		
Science Enhancement Programme			X		
Assessment for Learning (AfL)			X		
Dialogic teaching			X		
Puppet project			X		
Learning skills for science			X		
IPSE – primary advisory teachers			X		
AKSIS Project			X		
Concept cartoons			X		
Getting practical			X		
APP			X		
School maths project group				X	
AS Use of Maths				X	
Nuffield Projects (Biology, Chemistry, Physics, Combined)				X	
Nuffield KS3 STEM Project				X	
SATIS				X	
Suffolk Science				X	
Twenty-first Century Science				X	
Applied Science				X	
QCA Schemes of work				X	

Thinking through Geography				X	
Open Minds Project (RSA)				X	
Ginn Science				X	
Nuffield D&T				X	
Salters Science				X	
Wellcome Trust Darwin projects					X
Wellcome / Camden STEM Project					X
Reachout Lab Imperial College					X
Get ahead with STEM (Smallpiece Trust)					X
Crest Awards					X
Chemistry for our future					X
School labs based in universities					X
WISE					X
Children Challenging industry					X
HE STEM Programme					X
STEM Ambassadors					X
STEM Clubs					X
Planet science					X
Cre8te maths					X
Upd8 including primary					X
Nuffield bursary scheme					X

The future of science depends on the quality of science education today.

We are a global charitable foundation dedicated to achieving extraordinary improvements in human and animal health. We support the brightest minds in biomedical research and the medical humanities. Our breadth of support includes public engagement, education and the application of research to improve health. We are independent of both political and commercial interests.

Wellcome Trust
Gibbs Building
215 Euston Road
London NW1 2BE, UK
T +44 (0)20 7611 7221
F +44 (0)20 7611 8254
E education@wellcome.ac.uk
wellcome.ac.uk

The Wellcome Trust is a charity registered in England and Wales, no. 210183. Its sole trustee is The Wellcome Trust Limited, a company registered in England and Wales, no. 2711000 (whose registered office is at 215 Euston Road, London NW1 2BE, UK). MP-5991/5-2014/JC