REVIEW

August 2003

A systematic review of the effects of context-based and Science-Technology-Society (STS) approaches in the teaching of secondary science

This review is supported by the Teacher Training Agency (TTA) to promote the use of research and evidence to improve teaching and learning

Review conducted by the TTA-supported Science Review Group

The EPPI-Centre is part of the Social Science Research Unit, Institute of Education, University of London

REVIEW TEAM

Authors and review team

Dr Judith Bennett, University of York
Dr Sylvia Hogarth, University of York
Fred Lubben, University of York

Administrative support

Alison Robinson, University of York

EPPI-Centre advisors

Professor Diana Elbourne, EPPI-Centre, Institute of Education, London
Ms Carole Torgerson, EPPI-Centre and University of York

Contact details

Dr Judith Bennett
Department of Educational Studies
University of York
York YO10 5DD

tel: 01904 433471
e-mail: mailto:jmb20@york.ac.uk.

ACKNOWLEDGEMENTS

The Science Review Team is one of a number of teams undertaking systematic reviews to support the evidence base of Initial Teacher Training (ITT) staff in England.

The Review Team is grateful for the funding from the Teacher Training Agency (TTA) which enabled this review to be carried out. The Review Team also acknowledges the financial support via core institutional funding from the Higher Educational Funding Council for England (HEFCE). The Review Team is grateful to the EPPI-Centre at the Institute of Education in London for training and support for the review, and for permission to use their systematic review tools.

Possible conflicts of interest

Two members of the Review Team, Judith Bennett and Fred Lubben, have been involved in the development of school science materials which emphasise context-based approaches to the teaching of science. All members of the Review Team have been involved in evaluating context-based courses and have published in this area. In fact, one of the authors of the Review (Bennett) is the author of one of the five studies included in the in-depth review (Ramsden, 1997). Another study included in the in-depth review concerns a thesis undertaken under supervision of a colleague of the Review Team members. In addition, the Review Team members are members of the University of York Science Education Group (UYSEG), which is currently developing a new secondary level science course, *21st Century Science*, aimed at developing pupils' scientific literacy. However, none of the team members is directly involved in the development of this course.

LIST OF ABBREVIATIONS

BEI	British Education Index
DfES	Department for Education and Science
EPPI-Centre	Evidence for Policy and Practice Information and Co-ordinating Centre
ERIC	Education Resources Information Centre
GCSE	General Certificate for Secondary Education
ILL	Inter-Library Loan
ITT	Initial Teacher Training
PISA	Programme for International Student Assessment
PLON	Physics Curriculum Development Project (Dutch)
QCA	Qualifications and Curriculum Authority
RCT	Randomised Control Trial
SATIS	Science And Technology In Society
SAT	School Assessment Test
SS&C	Scope, Sequence & Continuity
SSCI	Social Science Citation Index
STS	Science, Technology and Society
TIMMS	Third International Mathematics and Science Survey
TTA	Teacher Training Agency
VOSTS	Views On Science Technology and Society

This report should be cited as: 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. Version 1.1 In: *Research Evidence in Education Library*. London: EPPI-Centre, Social Science Research Unit, Institute of Education.

TABLE OF CONTENTS

SUMMARY

Introduction and background

Underpinning the topic which has formed the focus of this review is a situation most teachers will recognise: many of their pupils simply do not seem to find science lessons particularly interesting or see the point of what they are doing. This concern is not new; there are examples of articles dating back to the 1920s in which very similar worries are expressed.

The last twenty years or so has seen a number of changes in science teaching, of which one of the most significant has been the development of a wide range of materials which use contexts and applications as a starting point for developing an understanding of scientific ideas. Such approaches are variously described as 'context-based', 'applications-led' or using 'STS' (Science-Technology-Society) links. Examples of curriculum development drawing on such approaches can be found in materials ranging from small teaching units to whole courses, developed on local, national and international scales, and for all age ranges from primary through to tertiary. A key aim of these approaches is to stimulate young people's interest in science and to help them see how it relates to their everyday lives.

Given the aspirations of these approaches and their widespread use, it is very important to examine their effects in a systematic way. This report therefore presents the work undertaken for a systematic review of **the effects on pupils of teaching approaches which emphasise placing science in context and promote links between science, technology and society (STS).**

The reasons for selecting this area for review are as follows:

- Courses adopting context-based and STS approaches have attracted national and international attention in the last two decades as they are seen to have an important role to play in developing pupils' scientific literacy.
- Approaches which set science in context and help pupils see links between science, technology and society are strongly advocated in the *National Curriculum for Science* in England and Wales, and in the *Qualifications and Curriculum Authority (QCA) Scheme of Work for Key Stage 3.*
- Approaches which set science in context and help pupils see links between science, technology and society have been advocated for a number of years on Initial Teacher Training (ITT) courses.
- It seems highly likely that the current move towards developing school science courses which emphasise scientific literacy will continue, and that this will have significant implications for science ITT courses. The overlap between the aims of context-based and STS courses and scientific literacy courses, together with the approaches they advocate, makes it highly desirable to establish the strength and nature of the evidence base for the claims made for such approaches.

Research into the effects of context-based and STS approaches falls into three main areas. The most significant concerns pupils' affective responses. A number of people working in science education, particularly those involved in the development of the associated curriculum materials, have argued that they are very motivating for pupils, and there is some evidence to indicate that pupils do respond positively in lessons where such approaches are used. A second strand of research has focused on the development of pupils' understanding of scientific ideas as a result of following context-based approaches. The final strand has explored aspects of teachers' responses to, and use of, materials incorporating context-based and STS approaches. Broadly speaking, the research appears to indicate that pupils following context-based and STS courses develop an understanding of scientific ideas which is at least as good as that of pupils following more conventional approaches, and that such approaches do appear to motivate pupils in lessons. One purpose of this review is to establish the strength of the evidence base which supports these claims.

Aims of the review and review question

The review has two principal aims:

- to explore the effects on pupils of teaching approaches which emphasise placing science in context and promote links between science, technology and society (STS)
- to inform the evidence base on which Initial Teacher Training (ITT) courses draw in relation to the above teaching approaches

The research review question is:

- What evidence is there that teaching approaches which emphasise placing science in context and promote links between science, technology and society improve the understanding of science ideas and the attitudes to science of 11 to 18-year-old pupils, and what are the implications of the evidence for initial teacher training courses?

The focus on understanding encompasses science concepts, ideas about the nature of science, and scientific method. The focus on attitude encompasses attitude towards science, attitude towards school science, motivation to learn, interest in science activities, and career intentions.

The mapping of the area revealed a wide range of relevant studies. A more limited focus was therefore adopted for the in-depth review, with the review question being refined to focus on the effects on pupils following whole context-based courses.

The question for the in-depth review is therefore:

- What evidence is there from controlled evaluation studies that context-based courses improve understanding of science ideas and the attitudes to science of 11 to 18-year-old pupils, and what are the implications of the evidence for initial teacher training courses?

Review methods

The review methods are those developed by the EPPI-Centre for systematic reviews of educational research literature. The review involves a systematic process which takes place in three main phases.

The phase first involves developing criteria by which studies are included or excluded, searching (though electronic databases and by hand) for studies which appear to meet these criteria, and then screening the studies to see if they meet the inclusion criteria. The second phase involves coding each of the included studies against a pre-agreed list of characteristics. This process is called keywording. Once studies have been keyworded, they are used to generate a systematic map of the area. This groups studies according to their chief characteristics.

The third phase involves selecting an area of the map for in-depth review. Once this area has been identified, the contents of each paper are summarised and evaluated according to pre-agreed categories. This process is called data-extraction. The extracted data are then pulled together to form a key section in the report.

Results

The number of studies identified through the searching and screening processes demonstrated that there are high levels of interest in the review topic, and application of inclusion and exclusion criteria identified 66 studies for inclusion in the systematic map.

The systematic map allowed a number of characteristics of the work to be identified. In particular, the map has shown the following:

- The majority of the work has taken place in the USA, the UK, the Netherlands and Canada.
- Pupils in the 11 to16 age range are seen as the target for the majority of the interventions using context-based or STS approaches.
- There are comparable levels of interest in the effects of such approaches on both understanding of ideas and attitudes to science or science lessons.
- A diversity of measures are used to assess effects on understanding and attitudes.
- 26 studies have drawn on designs which make use of control groups in order to strengthen any claims they make.

Five studies were included in the in-depth review, which focused on controlled evaluation studies of context-based courses which reported on both understanding and attitudes. The in-depth review demonstrated that research in the area is generally of a good quality and, as such, confidence can be placed in its findings.

The in-depth review has shown the following:

- There is some evidence to support the claim that context-based approaches motivate pupils in their science lessons.
- There is evidence to support the claim that such approaches also foster more positive attitudes to science more generally.
- There is good evidence to support the claim that context-based approaches do not adversely affect pupils' understanding of scientific ideas.

Conclusions

The first conclusion drawn from the review is that the process has been very worthwhile in pulling together in a systematic way the research evidence on a very important area in science education, particularly in the context of current developments in the school science curriculum.

Secondly, the in-depth review has indicated that research in the area is of a quality which allows reasonable confidence to be placed in the findings, suggesting that additional areas of the systematic map would benefit from in-depth review. These areas include interventions which focused on whole STS courses; and attitudes alone or understanding alone; and effects on disadvantaged pupils, pupils in specific ability ranges and pupils of different gender.

Thirdly, the systematic map has also identified areas where further research would be useful and informative, such as research into the effects of the more novel of teaching activities associated with context-based and STS approaches. Small group discussion work would be a case in point here.

Finally, the process of undertaking the review has helped the Review Team members clarify their thoughts in two key areas. The first of these concerns the ways in which research is reported, and what makes for clear, concise and accessible reporting of research studies. The second area concerns the nature of educational research and what characteristics it needs to have in order to be judged of good quality.

Recommendations

The Review Team is cautious about making recommendations for policy and practice on the basis of an in-depth review of only five studies, even though these have been identified through a rigorous series of procedures. However, the consistency of the evidence which emerged from the in-depth review does suggest that its findings should be made available in a suitable form to key groups, such as students on initial teacher training courses and their university- and school-based tutors. Other key groups for whom the findings are of relevance and importance are those who formulate policy for initial teacher training courses and who develop materials which students encounter on their courses. These include the Department for Education and Science (DfES), who are responsible for introducing the Key Stage 3 Strategy, and the Qualifications and Curriculum Authority (QCA), who have produced an extensive and detailed Scheme of Work for Key Stage 3.

The Review Team has noted in its report that there is increasing anecdotal evidence that many science lessons now begin with pupils copying down into their books the intended learning outcomes of the lesson from the board or an overhead projector transparency. Whilst it is certainly important that both initial teacher training students and pupils are clear about what they hope pupils will learn in their lessons, this is a necessary but not sufficient condition for high quality science teaching. It is clear from the review that context-based approaches provide an effective way to interest and motivate pupils in their science lessons.

One other recommendation the Review Team would like to make concerns dissemination of the findings of the review. The team supports the efforts made to summarise and disseminate the findings in formats appropriate for a variety of intended audiences, such as ITT practitioners and policy-makers.

1. BACKGROUND

1.1 Introduction

The last twenty years has seen a number of changes in science teaching, of which one of the most significant has been the development of a wide range of materials which use contexts and applications as a starting point for developing an understanding of scientific ideas. Such approaches are variously described as ‘context-based’, ‘applications-led’ or using ‘STS’ (Science-Technology-Society) links. Pupils using such materials might, for example, find out about the electromagnetic spectrum through learning about medical techniques for seeing inside the body, or explore the views of different members of a community on the impact of locating a chemical industry nearby. Examples of curriculum development drawing on such approaches can be found in materials ranging from small teaching units to whole courses, developed on local, national and international scales, and for all age ranges from primary through to tertiary.

Given the widespread use of these approaches, it is very important to examine their effects in a systematic way. This report therefore presents the work undertaken for a systematic review of **the effects on pupils of teaching approaches which emphasise placing science in context and promote links between science, technology and society (STS).**

The review topic falls within the broader area of *scientific literacy*, which is currently receiving considerable attention in policy-making discussions about the future of the school science curriculum. The reasons for selecting context-based approaches and approaches which promote links between science, technology and society are as follows:

- Courses adopting context-based and STS approaches have attracted national and international attention in the last two decades as they are seen to have an important role to play in developing pupils’ scientific literacy.
- Context-based and STS approaches are strongly advocated in the *National Curriculum for Science* in England and Wales, and in the *Qualifications and Curriculum Authority (QCA) Scheme of Work for Key Stage 3.*
- Context-based and STS approaches have been advocated for a number of years on ITT courses as a means by which teachers might motivate their pupils.
- It seems highly likely that the current move towards developing school science courses which emphasise scientific literacy will continue, and that this will have significant implications for science ITT courses. The overlap between the aims of context-based and STS courses and scientific literacy courses, together with the approaches they advocate, makes it highly desirable to establish the strength and nature of the evidence base for the claims made for such approaches.

1.2 Definitions

The following definitions have been adopted for the purposes of this review:

Context-based approaches are approaches adopted in science teaching in which contexts and applications of science are used as the *starting point* for the development of scientific ideas. This contrasts with more traditional approaches, which cover scientific ideas first, and conclude with a brief mention of applications.

Science, Technology and Society (STS) is a term which can be used very broadly. This review has adhered to the definition provided by Aikenhead (1994, pp 52-53), who defines STS approaches as those which emphasise links between science, technology and society by means of emphasising one or more of the following: a technological artefact, process or expertise; the interactions between technology and society; a societal issue related to science or technology; social science content that sheds light on a societal issue related to science and technology; a philosophical, historical, or social issue within the scientific or technological community.

Scientific literacy is a term frequently encountered in writing about context-based and STS approaches. In the context of this review, it has been taken to encompass the knowledge, understanding and skills young people need to develop in order to think and act appropriately on scientific matters which may affect their lives and the lives of other members of the local, national and global communities of which they are a part.

In making the above distinctions between context-based approaches and STS approaches, it is recognised that there are considerable areas of overlap in terms of the content of materials. However, it is possible to identify some general patterns of use, though with the proviso that there are exceptions. The term 'context-based' is more common in Europe, whilst 'STS' is more common in North America. 'Context-based' is also a term more frequently applied to whole courses and 'STS' to smaller units of enrichment materials. However, there are examples, particularly in the US of whole courses which are described as 'STS courses' and which share many of the characteristics of course described as 'context-based' elsewhere.

1.3 Policy and practice background

1.3.1 Interest in context-based and STS approaches

The origins of context-based approaches and their links with STS approaches have been documented by Bennett (2003) and Bennett and Holman (2003). Ideas underpinning the development of context-based approaches have also been documented by Campbell *et al.* (1994). Aikenhead (1994) has produced a detailed overview of STS approaches and materials, and how they draw on contexts and applications to develop ideas about science, technology and society.

The term 'context-based approaches' appears to have been applied to some of the activities in school science classrooms for a little under twenty years. Context-based approaches have their origins in two types of course: those for less academic pupils at the secondary level, such as *Nuffield Science 13-16* (Nuffield Foundation, 1980), and those for non-science specialists at the upper secondary or tertiary level, such as *Science In Society* (Lewis, 1981) in the UK, and *Science and Culture* (Yager and Casteel, 1968) in the US.

From the early 1980s, context-based approaches started to appear in mainstream science courses and these are now in widespread use in a number of countries. Examples of such materials tend to be found at the high school level and include, for example, in the UK, *Science: the Salters Approach* (University of York Science Education Group (UYSEG), 1990-1992), *Salters Advanced Chemistry* (Burton *et al.,* 1994) and the SLIP (Supported Learning in Physics Project, 1997) in the UK. Examples from other countries include *ChemCom* (American Chemical Society [ACS], 1988) in the US, *PLON* (Dutch Physics Curriculum Development Project, 1988) in the Netherlands; and *STEMS* (Science, Technology Environment in Modern Society) in Israel (Tal *et al.,* 2001).

Courses using context-based and STS approaches are normally characterised by one or both of the following aims: to help young people appreciate how science relates to their current and future lives, and to stimulate interest in science, possibly with a view to encouraging more young people to continue their study of science beyond the compulsory period. Additionally, they tend to be characterised by a broader range of teaching strategies than the traditional teacher exposition and practical work associated with science teaching. Thus context-based and STS approaches advocate the use of small-group discussions, role-plays, poster presentations, problem-solving tasks, creative writing and a variety of activities aimed at encouraging pupils to interact with material they are reading, such as Directed Activities Related to Text (DARTs) (Davies and Green, 1984).

1.3.2 Links with initial teacher training (ITT)

ITT courses have tended to promote the use of context-based and STS approaches for several reasons. Firstly, they are seen as motivating for pupils. Secondly, the range of teaching strategies they include gives them strong links with ‘pupil-centred learning’ or ‘active learning’. These terms are applied to activities where pupils have a significant degree of autonomy over the learning activity, and are frequently advocated in teaching as a means of stimulating pupils’ interest in what they are studying (e.g. Kyriacou, 1998). Thirdly, and more recently, context-based and STS approaches, with their associated emphasis on discussion and writing tasks, have been promoted as a means for developing pupils’ language skills in science, such as in the use of argument (e.g. Newton *et al.,* 1999; Osborne *et al.,* 2001) and the development of writing genres (e.g. Wray and Lewis, 1997).

1.4 Research background

Research into the effects of context-based and STS approaches falls into three main areas. The most significant concerns pupils’ affective responses. A number of people working in science education, particularly those involved in the development of the associated curriculum materials, have argued that they are very motivating for pupils (e.g. Bybee, 1985; Eijkelhof and Kortland, 1988; Fensham, 1988; Hofstein *et al*., 1988; Campbell *et al.,* 1994), and some evidence is claimed to support this notion (e.g. Ramsden, 1992). Within this, one area of focus has been to explore their effects in relation to gender, with context-based and STS approaches being seen as a means of encouraging more girls to be interested in science. A second strand of research has focused on the development of pupils’ understanding of scientific ideas as a result of following context-based and STS approaches (e.g. Barker and Millar, 2000). The final strand has explored aspects of teachers' responses to, and use of, materials incorporating context-based and STS approaches (e.g. Borgford, 1995). Broadly speaking, the claims made by the research are that pupils following context-based and STS courses develop an understanding of scientific ideas which is at least as good as that of pupils following more conventional courses, and that such approaches do appear to motivate pupils in lessons. One purpose of this review is to establish the strength of the evidence base on which these claims are made.

1.5 Authors, funders and other users of the review

1.5.1 Why the Science Review Team is an appropriate group to undertake the review

The members of the Review Team all have considerable research experience in science education, both in the UK and in a number of other countries. Members of the Review Team are also members of the University of York Science Education Group (UYSEG), a group with an international reputation for its research, curriculum development and teacher professional development.

One of the team, Judith Bennett, has recently published a book, *Teaching and learning science: a guide to recent research and its applications* (Bennett, 2003). As Chair of the Research Committee of the Association for Science Education (ASE), she is directly involved in the promotion of the use of research findings in initial teacher training (ITT) and continuing professional development (CPD) programmes. She also has a significant role in the Science ITT programme in the Department of Educational Studies at York, and was responsible for the introduction at York of a part-time research-based masters programme, the MA in Science Education, for practising teachers. A number of participants in this programme are teachers associated with the Department of Educational Studies' ITT Partnership, and/or former students on the ITT course. The Review Team has therefore an established track record of working closely with teachers and building research capacity in the area of the review.

The Review Team is also part of the EPPI-Centre Science Review Group, using similar review procedures as required for this review.

1.5.2 The topicality of the review

The publication of *Beyond 2000* (Millar and Osborne, 1998) stimulated debate over the nature of the school science curriculum and has resulted in moves to increase the emphasis on *scientific literacy*; in other words, to equip pupils with the knowledge, understanding and skills they need as future adult citizens in order that they can engage with discussion and debate of scientific matters and issues. This overlaps considerably with the aims of context-based and STS approaches. The recent report of the House of Commons Science and Technology Committee (House of Commons, 2002) highlights the prominent role of scientific literacy: 'A new National Curriculum should require all pupils to be taught the skills of scientific literacy and selected key ideas across the science' (p 5). Examples of courses specifically aimed at developing scientific literacy are now emerging, such as *Science for Public Understanding* (Hunt and Millar, 2000), a post-compulsory course for 17- to 18-year-olds, and *21st Century Science*, a GCSE course currently being developed by the University of York and the Nuffield Curriculum Centre.

1.5.3 The potential audiences and users of the findings of the review

The work done for the review, and its findings, are of interest to a number of groups. These include university staff and school staff involved in teaching and supervision on initial teacher training courses, the students on such courses, people who write textbooks for such courses and also for school science courses, policy-makers responsible for making decisions about the content of initial teacher training courses, those involved in the development of curriculum materials which emphasise context-based and STS approaches, and parents with an interest in the experiences being offered to their children in school science lessons.

1.6 Review question

The review has two principal aims:

- to explore the effects on pupils of teaching approaches which emphasise placing science in context and promoting links between science, technology and society (STS)
- to inform the evidence base on which Initial Teacher Training (ITT) courses draw in relation to the above teaching approaches.

The research review question is:

- What evidence is there that teaching approaches which emphasise placing science in context and promote links between science, technology and society improve the understanding of science ideas and the attitudes to science of 11 to 18-year-old pupils, and what are the implications of the evidence for initial teacher training courses?

The focus on understanding encompasses science concepts, ideas about the nature of science, and scientific method. The focus on attitude encompasses attitude towards science, attitude towards school science, motivation to learn, interest in science activities, and career intentions.

In undertaking the review, it became apparent that a number of studies made reference to other science-related abilities, such as the development of investigative skills, manipulative skills, communication skills, problem-solving skills and decision-making skills. Whilst this area was not the main focus of the review, such effects were noted when studies were categorised (keyworded) to produce an overview (the systematic map). It is recognised that, though some of these skills are not science-specific, there is an expectation that they will be applied in science contexts. It is further recognised that there are areas of debate over the interpretation of some of these terms (e.g. attitude, problem-solving skills) and how they might be measured.

In producing the systematic map of the area, it became apparent that the number of studies involved required a narrower focus to be taken for the in-depth review. It was therefore decided to focus the in-depth review on studies of effects on understanding and attitudes of pupils who followed *whole courses* using context-based approaches.

The review question for the in-depth review is therefore:

- What evidence is there from controlled evaluation studies that context-based courses improve understanding of science ideas and the attitudes to science of 11 to 18-year-old pupils, and what are the implications of the evidence for initial teacher training courses?

2. METHODS USED IN THE REVIEW

This review employed the systematic review methods developed by the Evidence for Policy and Practice Information and Co-ordinating Centre (the EPPI-Centre).

There are ten main steps in a systematic review of research literature. These are as follows:

- identifying the area for review
- formulating the research question for the review
- identifying criteria for including and excluding studies
- searching for studies which might be suitable for inclusion in the review
- screening the studies to see if they meet the inclusion criteria
- coding each of the included studies (keywording)
- characterising the included studies (systematic mapping)
- selecting an area of the map for in-depth review
- summarising and evaluating the contents of each of the studies included in the in-depth review (data-extraction)
- reporting on the work

The advantages of this strategy are as follows: the process is systematic and therefore trustworthy; it develops research capacity in terms of the products of each part of the process (studies identified, keyworded and data-extracted); and it develops research capacity in the Review Team.

The timescale specified for the review (11 weeks) and the level of resource meant that it was not possible to undertake a full systematic review, as these typically take between six and twelve months. This means that the following have been limited in the review:

- the extensiveness of the search strategy, particularly concerning unpublished studies
- the cut-off date for obtaining hard copies of studies
- the subset of studies included in the in-depth review

The reduction in studies synthesised appears to be preferable to making compromises on other aspects of the quality of the systematic review process.

2.1 User involvement

The timescale for the review precluded formal user involvement in the identification of the review question and the review process. Informally, conversations were held with other staff in the University of York Science Education Group (UYSEG) and colleagues involved in other networks of which the Review Team are members, such as the Research Committee of the Association for Science Education. These conversations suggested that a systematic review of an area with direct relevance to scientific literacy was highly desirable.

2.2 Identifying and describing studies to produce a map of the area

2.2.1 Defining relevant studies: inclusion and exclusion criteria

The criteria for including studies in the review are given below, followed by amplification where appropriate.

Studies have been *included* in the systematic map if:

1. they concern courses which teach science through contexts and applications or if they are about teaching courses which emphasis links between science, technology and society;
2. they address aspects of pupils' understanding in science or they address aspects of pupils' attitudes to science;
3. they are empirical studies of evaluations;
4. they concern pupils in the 11 to 18 age range;
5. they have been published (or unpublished but in the public domain) in the period 1980-2002;
6. they are published in English.

The review has focused on school age pupils because one of its aims is to inform the evidence base on which initial teacher training courses draw. As the majority of the mainstream curriculum development work emphasising the teaching of science in context and links between science, technology and society has been for the 11 to 18 age range, this is the most appropriate target for the review.

The review covers the period 1980-1992 because mainstream context-based and STS courses have their origins in the science curriculum development work of the early 1980s. Although there are likely to be important studies on context-based/STS learning in Southern America, the limited timescale precluded including studies in languages other than English.

Studies have been *excluded* from the systematic map if they are not about science, not about relevant aspects of science, not of specified study type, not on the specified age range, and not within the specified period.

The exclusion criteria are therefore as follows:
1. Exclusion 1: exclusion on topic (not science)
2. Exclusion 2: exclusion on aspects of science (not science in context or STS)
3. Exclusion 3: exclusion on learning effects (not understanding or attitude)
4. Exclusion 4: age (not 11 to 18)
5. Exclusion 5: exclusion on study type:
    (a) editorials, commentaries, book reviews
    (b) policy documents
    (c) resources
    (d) bibliography
    (e) theoretical (non-empirical) papers
    (f) methodology papers
    (g) review papers
6. Exclusion 6: language (not English)
7. Exclusion 7: date (not 1980-2002)

Appendix 2.1 gives the sheet used when applying the inclusion and exclusion criteria.

2.2.2 Identification of potential studies: search strategy

The search strategy for identifying potential studies to be included in the review involved exploring a number of sources, as described below:

- a hierarchy of bibliographic databases in the following order: ERIC, BEI, PsycInfo and SSCI (Appendix 2.2 gives details of the search terms used)
- handsearching of *Studies in Science Education*, a journal which specialises in major review articles in science education in order to identify in bibliographies any studies of potentially useful studies
- personal contacts
- direct requests to key informants, following completion of electronic and handsearching

Searches of these sources were limited so as to identify studies conducted in the time period 1980-2002.

2.2.3 Screening studies: applying the inclusion and exclusion criteria

The Review Team set up a database system (using *EndNote* software) for keeping track of, and coding, studies found during the review. Titles and abstracts were imported or entered manually into the review database. Inclusion and exclusion criteria were applied to titles and abstracts. Studies excluded on the basis of titles and abstracts were recorded on the database, together with reasons for their exclusion. Excluded studies of potential interest for theoretical and policy background were marked as such. Full reports of potentially relevant studies were sent for through interlibrary lending (ILL). All but two of the papers obtainable arrived by the cut-off deadline. Inclusion and exclusion criteria were re-applied to the full reports and those which did not meet the initial criteria were excluded. The database was fully annotated with reviewer decisions on inclusion and exclusion together with reasons for exclusion. At various stages, duplicate studies were removed and stored separately.

The process of searching and screening yielded 66 studies for inclusion in the review.

2.2.4 Characterising the included studies

Studies identified for inclusion in the review were keyworded. The purpose of this process is to describe the characteristics of studies, as opposed to summarising and judging the quality of the findings.

Keywording involves applying a standard sheet, the *EPPI-Centre Core Keywording Sheet* (EPPI-Centre, 2002a) supported by the *EPPI-Centre Core Keywording Strategy* (EPPI-Centre, 2002b). This sheet is given in Appendix 2.3 and is used to keyword all systematic reviews in education which use EPPI-Centre methods.

It needs to be noted that this review used a particular interpretation of evaluations, one of the study types included in category 10 of the core keywording sheet (Appendix 2.3) The EPPI-Centre Guidelines consider two different sorts of evaluation: (1) where the intervention being evaluated occurs as a normal part of practice which may then be evaluated *post hoc* ('naturally occurring') and (2) where the intervention is introduced specifically in order to be evaluated, i.e. solely as part of research ('researcher-manipulated'). Both types of evaluations may or may not have control groups outside the intervention. For this review, however, evaluations where a researcher actively allocates pupils to different learning approaches for the purposes of the research seemed unrealistic for large-scale curriculum interventions, but the issue of whether or not there was a control group was considered particularly relevant. Hence, 'evaluations of researcher-manipulated interventions' was re-defined for the purposes of this review to mean studies which focus on pupils taking existing programmes with a different learning approach, and with a control group as part of the research design. In contrast, the term 'evaluations of naturally occurring interventions' referred to evaluations in which there was no control group.

The EPPI-Centre core keywording sheet was supplemented by a review-specific keywording sheet, developed to characterise aspects of studies central to the main review question. For example, studies are differentiated according to the branch of science involved, the aspect of science learning on which the study focused, and the ways in which the outcomes were measured. This sheet is given in Appendix 2.3.

The data gathered through keywording were recorded on the electronic database. Once studies had been keyworded, they were scrutinised using the electronic database in order to produce the systematic map.

The following areas were scrutinised:

- linked reports
- country of study
- age/level of the pupils
- type of study
- discipline of the study
- nature of the intervention
- aspects of science learning the study was assessing
- outcome measures
- outcomes

The systematic map is presented in Figure 3.1 in section 3.1.

2.2.5 Identifying and describing studies: quality assurance processes

The application of inclusion and exclusion criteria was initially conducted by all three team members for a 2.5 percent random sample (58 studies). This was done independently in the first instance, and the team members then met to compare the codes allocated, discuss the discrepancies, and reach a consensus on how criteria were to be interpreted and applied. This enabled the clarification and removal of any ambiguities in their perceptions of coverage of the criteria. Only minor revisions were found to be necessary. Two team members were responsible for screening the remaining studies on the basis of abstract and title.

These team members and a member of the EPPI-Centre worked on a second 2.5% random sample (58 studies), working independently. These data were used to calculate inter-screener agreement using frequency counts and the Cohen's Kappa inter-screener reliability coefficients (see section 3.3).

Once the 66 studies to be included in the review had been identified, a 10 percent sample (six studies) was keyworded by all three team members to check the appropriateness of the review-specific keywords and reach a consensus on how keywords were to be applied. Again, the team first worked independently and then met to compare keywording, discuss the discrepancies and potential changes to the review-specific keywords, and reach a consensus on how keywords were to be interpreted and applied. Following minor modifications to the review-specific keywords, thus increasing the validity, a 29 percent sample (16 studies) was double-keyworded. A member of the EPPI-Centre also keyworded studies for a random sample of nine studies.

Once the five studies to be included in the review had been identified, a further check was undertaken by two team members on these studies to ensure that keywording had been done consistently and accurately. Other studies which appeared to come close to the inclusion criteria were also double-checked to ensure that appropriate decisions had been made.

2.3 In-depth review

2.3.1 Moving from broad characterisation (mapping) to in-depth review

The purpose of in-depth reviewing is to describe the characteristics of studies in more detail and to assess the quality of methods used and the findings of studies. An in-depth review involves summarising and evaluating the contents of each of the studies included in the in-depth review (data-extraction).

The review question was refined for the in-depth review in the light of what emerged in the systematic map.

The inclusion criteria for studies in the in-depth review are as follows:

- The studies are evaluations with a control group, i.e. 'researcher-manipulated'.
- The studies focus on pupils following whole courses which used context-based approaches.
- The studies report outcomes in relation to both understanding of science and attitudes to science.

For the purposes of this review, a course was defined as a coherent context-based, learning programme with an explicit curriculum outline which is subject to an accredited examining system. In contrast, units or modules of learning which are inserted into other programmes, as identified through the review-specific keywording sheet in Appendix 2.3, were excluded.

Thus studies were excluded from the in-depth review on the following basis:

1. Exclusion 1: exclusion on study type (study is not an evaluation with a control group, i.e. 'researcher-manipulated')
2. Exclusion 2: exclusion on focus (study is not about whole courses using context-based approaches)
3. Exclusion 3: exclusion on outcomes (study is not reporting findings on both understanding and attitudes)

2.3.2 Detailed description of studies in the in-depth review

Studies identified as meeting the inclusion criteria for in-depth review were double data-extracted and quality assessed, using the EPPI-Centre's detailed data-extraction software, EPPI Reviewer (EPPI-Centre, 2002c).

2.3.3 Assessing the quality of studies and weight of evidence for the review question

Once data have been extracted from the studies, the next step in the review is to assess the quality of the studies and the weight of evidence they present in relation to the review question. The EPPI-Centre data-extraction procedures identify three categories to help in the process of apportioning different weights to the findings and conclusions of different studies, with each of these being rated as high, medium or low.

The categories are as follows:

Category A: the trustworthiness of findings (internal methodological coherence) in relation to the study question

Category B: the appropriateness of the research design and analysis used for answering the review question

Category C: the relevance of the study topic focus (from the sample, measures, scenario, or other indicator of the focus of the study) to the review question

Finally, an overall weighting (Category D) is compiled based on the judgements reached in categories A, B and C above.

The nature of the review process meant that studies included in the in-depth review were unlikely to gain anything other than ratings of 'high' in Category C. Thus an algorithm was developed to assess the overall weighting for Category D which gave Category C a lower contribution to the overall weighting. Within categories A and B, there was a need to refine the ratings at the top scales in order to increase the discrimination. This resulted in a non-linear scale for these categories.

The algorithm for moving from categorisation of A, B and C to the overall weighting, D is as follows:

In Category A, studies were allocated the following numerical ratings:

- high = 6
- medium high = 4
- medium = 2
- medium low = 1
- low = 0

In Category B, studies were allocated the following numerical ratings:
- high = 6
- medium high = 4
- medium = 2
- medium low = 1
- low = 0

In Category C, studies were allocated the following numerical ratings:
- high = 2
- medium = 1
- low = 0

Category D was based on adding the scores in A, B and C. Studies were then allocated on the following basis:
- high = 12-14
- medium high = 9-11
- medium = 6-8
- medium low = 3-5
- low = 0-2

2.3.4 Synthesis of evidence

The final step in the review is to synthesise the findings and bring together the studies which answer the review questions and which meet the quality criteria relating to appropriateness and methodology.

2.3.5 In-depth review: quality assurance process

A number of steps were followed in the review process for the purposes of quality assurance. All three team members undertook a data-extraction on one of the in-depth review studies, working first individually and then meeting to moderate their summaries. This process increased the reliability of the subsequent data-extractions. Data-extraction was then conducted by pairs of team members working first independently and then comparing their decisions and coming to a consensus. The study by Ramsden (1997) authored by one of the members of the Review Team was data-extracted by all three team members. In addition, for purposes of quality assurance, two members of the EPPI-Centre double data-extracted and quality assessed three of the studies included in the in-depth review.

All three members of the Review Team discussed and agreed the final decisions about weightings. Part of these discussions involved the consistency of the application of criteria for judgements of quality across the in-depth review as a whole. Another key aspect of this moderation involved developing the algorithm described in section 2.3.3 to assign weightings to studies. The purpose of the algorithm was to give appropriate weightings to studies in each of the categories and to discriminate adequately between papers. The validity of the algorithm was checked by each team member independently assigning a ranking to the five studies, and these then being compared.

Additionally, the Review Team was interested in checking the validity of the electronic searching process. This was done by identifying from personal knowledge a number of key studies (31 in total) which it was felt should emerge in the searching process.

3. IDENTIFICATION AND DESCRIPTION OF STUDIES

3.1 Studies included from searching and screening

Sixty-six studies were included in the review after searching and screening had taken place. Figure 3.1 shows how these studies were identified from searching to map to synthesis.

Figure 3.2 presents the origin of each of the 66 studies included in the map. The hierarchy of electronic databases described in section 2.2.2 was used in compiling the electronic database entries in this table: that is, ERIC entries were imported first, followed by BEI entries, then entries from PsycInfo, and finally entries from SSCI. Any duplicated entries were removed from the electronic database of review papers. Thus, if a study appeared as an entry in ERIC, and again in BEI, its source is given as ERIC.

Figure 3.2: The sources of the 66 studies included in the systematic map

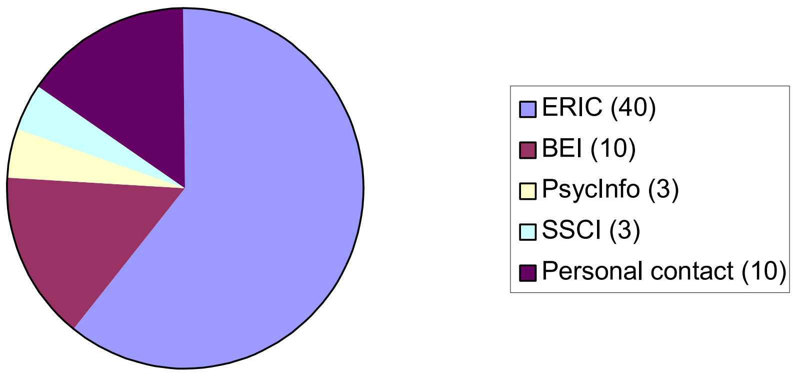
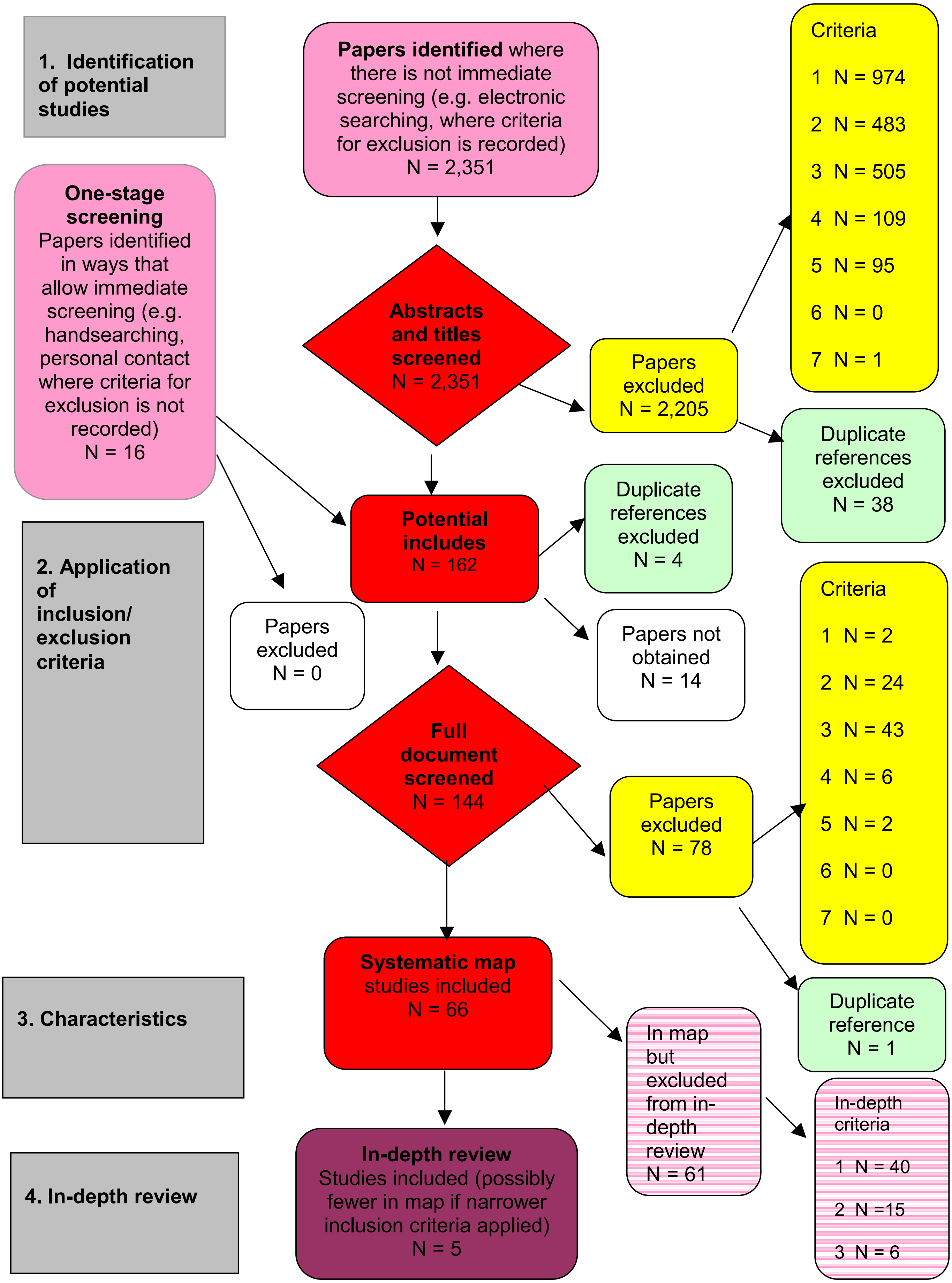


Figure 3.1: Filtering of papers from searching to map to synthesis



3.2 Characteristics of the included studies: the systematic map

As described in section 2.2.4, the following areas were scrutinised in detail to produce the systematic map:

- linked reports
- country of study
- age/level of the pupils
- type of study
- discipline of the study
- nature of the intervention
- aspects of science learning the study was assessing
- outcome measures
- outcomes

Of the 66 studies, 62 were published studies, two were unpublished MA theses, one was an unpublished PhD thesis and one was a set of conference proceedings. The nature of the review question meant that there was uniformity in responses to a number of the keywording categories. All the studies focused on curriculum and all focused on learners. Within this group, 12 also focused on teachers as well as learners. Two studies did not report a specific age range. The majority of the studies were undertaken in mixed-sex educational settings, with four taking place in single-sex settings, all of which were girls' schools.

3.2.1 Linked reports

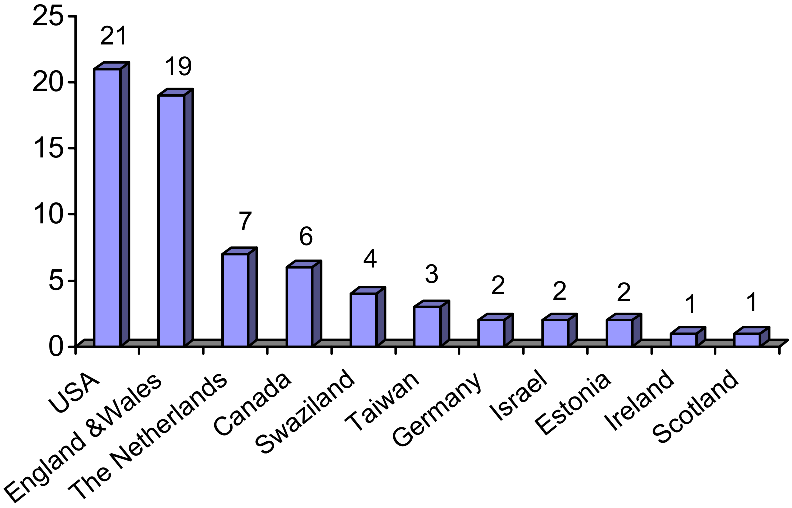
There were a number of examples of linked studies. These give some indication of clusters of activity and, in the first four listed below, point to co-ordinated research programmes. In particular, there were clusters of reports of studies undertaken by:

- researchers at the University of Utrecht in the Netherlands on the Dutch Physics Curriculum Project (PLON) (three papers)
- researchers at the University of York Science Education Group on the Salters curriculum development projects (eight papers) and context-based materials developed for use in Southern Africa (three papers)
- Yager and co-workers in Iowa the USA on a variety of aspects of the use of STS materials (two studies)
- Zoller and co-workers in Israel and British Columbia in Canada on STS courses for upper high school students (four studies)
- evaluators of the Science and Technology in Society (SATIS) materials (three papers)
- Solomon in the UK on the discussion of socio-scientific issues (three papers)

3.2.2 The country of study

Figure 3.3 shows the countries in which the studies were undertaken. The total number of countries is 68, as two studies were carried out in more than one country. Just over 80 percent were carried out in the US, the UK, the Netherlands and Canada.

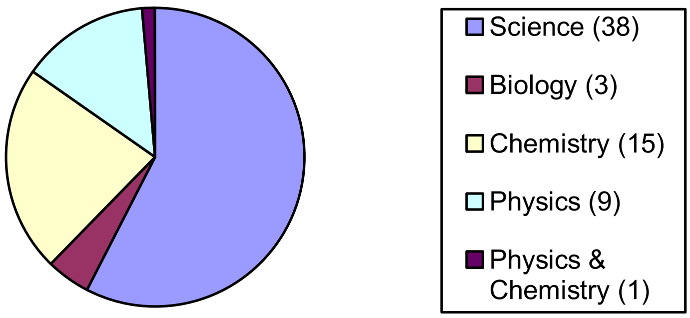
Figure 3.3: The country of study (N=66, not mutually exclusive, total=68)



3.2.3 The discipline of the study

Figure 3.4 shows the science discipline of the study.

Figure 3.4: The science discipline of study (N=66, mutually exclusive)



It is likely that the focus on chemistry and physics in the individual science disciplines in part reflects the motives for developing context-based materials in the first instance, with chemistry and physics being seen as subjects with a lower appeal than biology.

3.2.4 The age/level of the pupils

Forty-two studies were undertaken with pupils in the 11 to 16 age range, and 22 with pupils in the 17 to 20 age range. Within this, three studies had also gathered data from the 6 to 10 age range, and one from the age range 21 and over. Two studies did not give details of the age range of the pupils. The emphasis on pupils in the 11 to 16 age range is likely to reflect the perception of this age group as being very critical in terms of interest in science declining.

3.2.5 The type of study

Twenty-six of the studies were evaluations of researcher-manipulated policies or practices. Forty of the studies were evaluations of naturally occurring policy and practices.

3.2.6 The nature of the intervention

Figure 3.5 shows the nature of the intervention described in each of the studies.

Figure 3.5: The nature of the intervention (N=66, mutually exclusive)

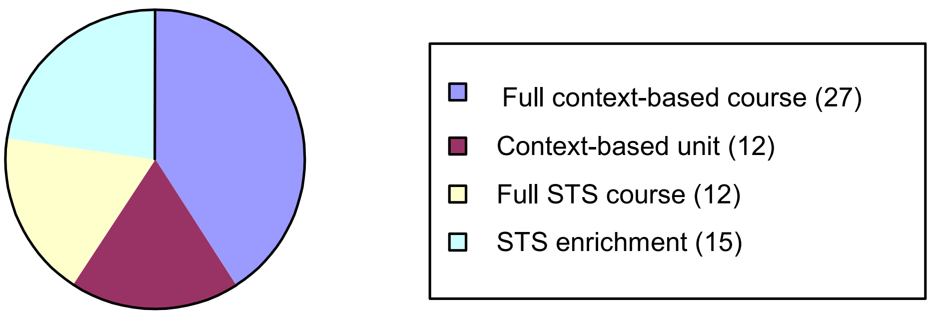


Figure 3.5 shows that slightly less than 60 percent (39 out of 66) of the studies evaluated context-based interventions, with more than two out of three of these dealing with full context-based courses. The remaining 40 percent of the studies report on evaluations of STS approaches. By way of contrast with context-based interventions, the majority of evaluations of STS approaches focus on STS enrichment activities within traditional approaches.

Table 3.1: The nature of the intervention in relation to the age range (N=66, mutually exclusive)

Nature of the intervention	Total	11 to 16 Age group (N=42)	17 to 18 Age group (N=22)	Age group not stated (N=2)
Full context-based course	27	14	13	-
Context-based unit	12	11	-	1
Full STS course	12	7	5	-

Nature of the intervention	Total	11 to 16 Age group (N=42)	17 to 18 Age group (N=22)	Age group not stated (N=2)
STS enrichment	15	10	4	1
Total	***66***	***42***	***22***	***2***

Table 3.1 indicates that studies evaluating the effect of full courses are over-represented amongst those dealing with the curriculum for the 17 to 18 age group. This applies to evaluations of full context-based and full STS courses, and may be explained by the interest in relationship of the teaching approach with the curriculum final assessment. In contrast, studies on smaller scale interventions, such as the introduction of context-based units and STS enrichment, are over-represented, more frequently focusing on the 11 to 16 age group. The outcomes in these cases may be intended to influence classroom teaching more directly.

Table 3.2: The nature of the intervention in relation to the nature of evaluation (N=66, mutually exclusive)

Nature of the intervention	Total	Researcher-manipulated (N=26)	Naturally occurring (N=40)
Full context-based course	27	11	16
Context-based unit	12	2	10
Full STS course	12	8	4
STS enrichment	15	5	10
Total	***66***	***26***	***40***

Table 3.2 suggests that evaluations of full STS courses are relatively more frequently researcher-manipulated in nature, whereas evaluations of the two types of small-scale interventions use relatively more frequently a naturally-occurring design. Studies of full context-based courses do not have a preference for either of the two evaluation designs.

3.2.7 The outcome measures

Figure 3.6 shows the outcome measures used in the studies. Figures total more than 66 as several studies employed a variety of different outcome measures.

Figure 3.6: Outcome measures (N=66, not mutually exclusive, total=123)

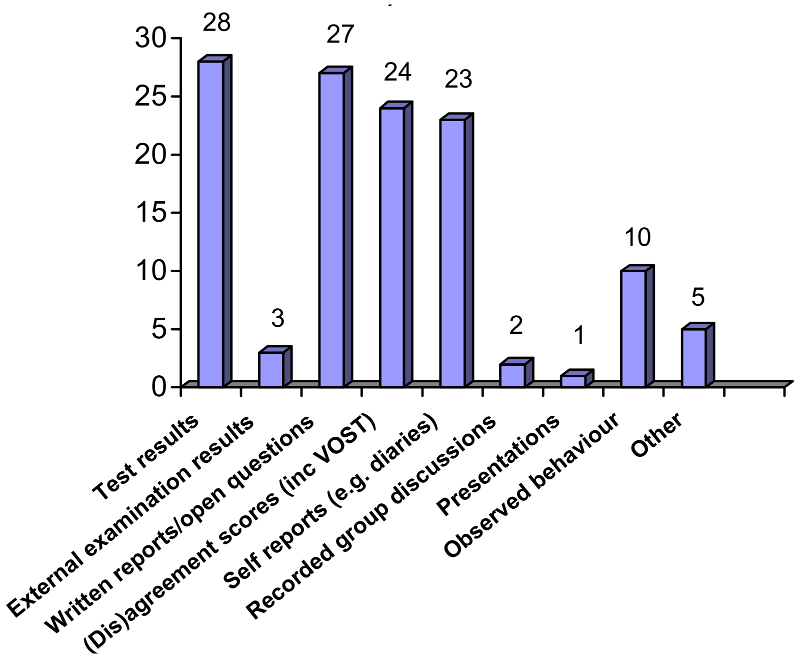


Figure 3.6 shows that about two in five of the studies used tests or written reports. Just over one in three of the studies used (dis)agreement scores, including Views on Science-Technology-Society (VOSTS) or self-reports.

Table 3.3: Outcome measures in relation to nature of evaluation (N=66, not mutually exclusive)

Outcome measures	Total	Researcher-manipulated (N=26)	Naturally occurring (N=40)
Test results	28	15	13
External examination results	3	-	3
Written reports/open questionnaires	27	11	16
Concept webs	-	-	-
(Dis)agreement scores (including VOSTS)	24	12	12
Self-reports (e.g. diaries, interviews)	23	5	18
Recorded group discussions	2	1	1
Presentations	1	-	1
Observed behaviour	10	3	7
Other	5	4	1
Totals	***123***	***51***	***72***

Table 3.3 shows the outcome measures as used in researcher-manipulated and naturally occurring evaluations. Of the four most common measures, only self reports (most usually interviews) emerged as being used much more frequently in

naturally-occurring evaluations than in researcher-manipulated evaluations; 18 out of 40 (around 45%) as compared with 5 out of 26 (around 20%). The reverse was the case for the use of test results, where over half (15 out of 26) of the researcher-manipulated interventions used this measure, compared with fewer than one- third (13 out of 40) of the naturally-occurring interventions.

3.2.8 The outcomes

Forty-six of the studies reported on attitudes and 44 on understanding. Of these, 24 reported on both these aspects. Two other aspects which also emerged as featuring prominently in studies were the development of skills (21 studies), and effects in relation to gender (20 studies).

Having looked at individual categories on the map, it was decided to undertake a cross-tabulation of selected categories to help with the process of identifying the studies for the in-depth review.

The cross-tabulation focused on the following characteristics:

- nature of the intervention
- age range
- science focus
- name of the course (to establish if any courses were receiving particular attention in research studies)
- nature of the outcomes reported

In relation to this last characteristic, in addition to noting outcomes in relation to attitude and understanding, it was decided to note outcomes in relation to skills and to gender, as these had emerged as aspects featuring prominently in a number of studies.

Tables 3.4 and 3.5 summarise these data. A comparison of these tables shows that studies on context-based/STS approaches in the different science disciplines are equally spread over the researcher-manipulated and naturally-occurring evaluations. Within each of these clusters, two in three studies focus on a science curriculum (16 out of 26 and 23 out of 40 studies), one in five focus on chemistry (six out of 26 and 9 out of 40) and one in seven focus on physics (three out of 26 and six out of 40).

Similarly, studies focusing on the two age-ranges are also relatively equally distributed over the two types of evaluations. Studies on the 11 to 16 age range constitute just over 60% of the researcher-manipulated and naturally-occurring evaluations (i.e. 18 out of 26 and 24 out of 40 studies, respectively). The remaining 40 percent of each type of evaluation focus on the 17 to 18 age range.

Table 3.4: Researcher-manipulated evaluations: outcomes measured
Total number of studies = 26 (11 to 16 age group N=18; 17 to 18 age group N=8)

Nature of interaction	Age range	Paper	Outcomes measured				Science	Course
			Attitude	**Understanding**	**Skills**	**Gender**		
Full context N = 11	11 to 16 N = 6	Gallagher *et al.* (1992)	✓		✓		Science	
		Ramsden (1997)	✓	✓		✓	Science	Science: The Salters Approach
		Smith and Bitner (1993)		✓	✓	✓	Chemistry	ChemCom
		Wierstra (1984)	✓	✓		✓	Physics	PLON
		Wiestra and Wubbels (1994)	✓	✓		✓	Physics	PLON
		Yager and Weld (1999)	✓	✓		✓	Science	Scope, Sequence and Continuity
	17 to 18 N = 5	Banks (1997)		✓			Chemistry	Salters Advanced Chemistry
		Barber (2000)	✓	✓	✓		Chemistry	Salters Advanced Chemistry
		Barker and Millar (1996)		✓			Chemistry	Salters Advanced Chemistry
		Key (1998)	✓				Chemistry	Salters Advanced Chemistry
		Winther and Volk (1994)		✓			Chemistry	Salters Advanced Chemistry
Context-based unit	11 to 16 N = 2	Dahncke *et al.* (2001)		✓	✓		Physics	
		Lubben *et al.* (1997)		✓	✓	✓	Science	Matsapha materials
Full STS courses N = 8	11 to 16 N = 5	Ben-Zvi (1999)	✓	✓			Science	Science and Technology for all
		Smith and Matthews (2000)	✓			✓	Science	
		Tsai (2000)		✓			Science	
		Yager *et al.* (1988)	✓	✓	✓		Science	
		Zoller *et al.* (1990)		✓		✓	Science	ST 11
	17 to 18	Solomon (1992)	✓		✓		STS	Discussions of Issues in School Science (DISS)
		Zoller *et al.* (1988)	✓			✓	Science	ST 11

Nature of interaction	Age range	Paper	Outcomes measured				Science	Course
			Attitude	Understanding	Skills	Gender		
	N = 3	Zoller *et al.* (1991)	✓				Science	ST 11
STS enrichment N = 5	11 - 16 N = 5	Christie and Nelson (1988)	✓				Science	
		Rubba (1990)	✓		✓		Science	
		Rubba *et al.* (1991)	✓	✓			Science	Genetics unit
		Tsai (1999)		✓			Science	
		Wiesenmeyer and Rubba (1999)	✓		✓		Sciences	
Context-based unit N = 10	11 - 16 N = 9	Campbell *et al.* (2000b)		✓	✓		Science	Linking School Science with Industry and Technology
		Dlamini *et al.* (1996)	✓			✓	Science	Matsapha
		Helms (1998)	✓				Science	Wetlands Restoration
		Kortland (1997)		✓			Physics	Garbage reusing/recycling
		Kortland (2001)		✓	✓		Science	
		Lubben *et al.* (1996)	✓	✓			Science	Matsapha
		Monhardt and Monhardt (1998)		✓	✓		Science	Jurisprudential Inquiry model to STS
		Roth (2001)		✓			Science	
		van Weelie (2001)		✓	✓		Science	Bio-diversity
	No age	Fowler and Sinclair (1992)	✓	✓	✓		Science	
Full STS courses N = 4	11 - 16	Ebenezer and Zoller (1993)	✓			✓	STS	ST 11
		Wiley (1991	✓				STS	
	17 - 18	Solomon (1988)		✓	✓		STS	
		Solomon (1989)	✓		✓		STS	DISS
		Ferreira (2001)	✓				Physics	

Nature of interaction	Age range	Paper	Outcomes measured				Science	Course
			Attitude	**Understanding**	**Skills**	**Gender**		
STS enrichment N = 10	11 to 16 N = 5	Rannikmäe (2001)	✓	✓	✓		Science	
		Stevens (2001)	✓				Science	
		Walker (1990)	✓				Science	SATIS
		Yager and Tamir (1993)	✓	✓	✓	✓	Science	Iowa STS Model
	17 to 18 N = 4	Alsop *et al.* (1998)		✓			Physics	
		Lenton (1990)	✓	✓		✓	Science	SATIS
		Lenton (1991)	✓	✓			Science	SATIS
		Phillips and Norris (1999)		✓			Science	
	No age N = 1	Williams *et al.* (1990)	✓	✓			Science/ Chemistry	The Illinois River Project

Table 3.5: Naturally-occurring evaluations: outcomes measured
Total number of studies = 40 (11 to 16 age group N=24; 17 to 18 age group N=14; age group not stated N=2)

Nature of intervention	**Age range**	**Paper**	**Outcomes reported**				**Science**	**Course**
			Attitude	Understanding	Skills	Gender		
Full context N = 16	11 to 16 N = 8	Eijkelhof and Lijnse (1988)	✓	✓	✓	✓	Physics	PLON
		Greenlee and Lambert (1996)	✓			✓	Science	The Science of Living Spaces
		Huppert *et al.* (1992)	✓	✓			Biology	Human Health and Science
		Licht (1990)	✓				Physics	PLON
		Nentwig *et al.* (2002)	✓				Chemistry	Chemie im Kontext
		Ramsden (1992)	✓	✓		✓	Science	Science: The Salters Approach
		Reid and Skryabina (2002)	✓			✓	Physics	Standard Grade Applications-led course
		Sutman and Bruce (1992)	✓	✓			Chemistry	ChemCom
	17 to 18 N = 8	Barker and Millar (1999)		✓			Chemistry	Salters Advanced Chemistry
		Barker and Millar (2000)		✓			Chemistry	Salters Advanced Chemistry
		Campbell *et al.* (2000a)	✓	✓			Physics	Salters Horners Advanced Physics
		Hughes (2000a)	✓	✓		✓	Chemistry	Salters Advanced Chemistry
		Hughes (2000b)	✓			✓	Chemistry	Salters Advanced Chemistry
		Pedersen (1992)	✓	✓			Chemistry	ChemCom
		Pilling *et al.* (2001)	✓	✓	✓	✓	Chemistry	Salters Advanced Chemistry
		Truex (1987)	✓	✓	✓		Science	Contemporary Science
Context-based unit N = 10	11 to 16 N = 9	Campbell *et al.* (2000b)		✓	✓		Science	Linking School Science with Industry and Technology
		Dlamini *et al.* (1996)	✓			✓	Science	Matsapha
		Helms (1998)	✓				Science	Wetlands Restoration
		Kortland (1997)		✓			Physics	Garbage reusing/recycling
		Kortland (2001)		✓	✓		Science	

Nature of intervention	**Age range**	**Paper**	**Outcomes reported**				**Science**	**Course**
			Attitude	Understanding	Skills	Gender		
		Lubben *et al*. (1996)	✓	✓			Science	Matsapha
		Monhardt and Monhardt (1998)		✓	✓		Science	Jurisprudential Inquiry model to STS
		Roth (2001)		✓			Science	
		van Weelie (2001)		✓	✓		Science	Biodiversity
	No age N = 1	Fowler and Sinclair (1992)	✓	✓	✓		Science	
Full STS courses N = 4	11 to 16 N = 2	Ebenezer and Zoller (1993)	✓			✓	STS	ST 11
		Wiley (1991)	✓				STS	
	17 to 18 N = 2	Solomon (1988)		✓	✓		STS	
		Solomon (1989)	✓		✓		STS	DISS
STS enrichment N = 10	11 - 16 N = 5	Ferreira (2001)	✓				Physics	
		Rannikmäe (2001)	✓	✓	✓		Science	
		Stevens (2001)	✓				Science	
		Walker (1990)	✓				Science	SATIS
		Yager and Tamir (1993)	✓	✓	✓	✓	Science	Iowa STS Model
	17 - 18 N = 4	Alsop *et al.* (1998)		✓			Physics	
		Lenton (1990)	✓	✓		✓	Science	SATIS
		Lenton (1991)	✓	✓			Science	SATIS
		Phillips and Norris (1999)		✓			Science	
	No age N = 1	Williams *et al*. (1990)	✓	✓			Science/ Chemistry	The Illinois River Project

3.3 Identifying and describing studies: quality assurance results

The quality assurance processes for screening and keywording described in section 2.2.5 were followed, with the following results.

Before the start of the search, 31 key studies were identified by the Review Team, including eight background studies. The validity of the search strategy was confirmed when the search strings for ERIC and BEI together produced 24 of these studies. The remaining seven publications included three theses, one conference paper, one book chapter, one obscure journal publication and one from the *Journal of Chemical Education* that had not, surprisingly, been entered onto any of the databases searched. All these seven publications were included in the search through the list of 'papers through contacts'.

All 2,351 studies resulting from the search were screened on title and abstract between two team members. The reliability of the screening was established by independent screening of a random 2.5 percent sample (58 studies) by two team members and an EPPI-Centre member. The inter-screener reliability as measured by the frequency method and the Cohen's Kappa method is shown in the Table 3.6. The Cohen's Kappa method has the advantage of compensating for chance agreement.

Table 3.6: Inter-screener agreement for first screening

	Frequency method		**Cohen's Kappa method**	
	Identical decisions	Inter-screener agreement	Cohen's Kappa coefficient	Inter-screener agreement
Screener 1 and Screener 2	53	91%	0.403	Fair
Screener 1 and EPPI member	51	88%	0.524	Moderate
Screener 2 and EPPI member	48	83%	0.225	Fair

The screening decisions show a consistently high agreement for the random sample (around 90%), especially inter-screener agreement for the subject specialist team members. Apart from the difference in subject specialism, an additional explanation for the somewhat lower agreement with the EPPI-Centre link person may be the fact that the latter was not involved in the first stage of moderation of the screening process.

The high reliability of the screening as shown by the frequency method contrasts with the fair-moderate agreement according to the Cohen's Kappa method. This pattern may be due to the fact that the review used a very wide search strategy, resulting in the need for excluding a very high proportion of studies at the screening stage. Such a skewed sample increases the chance agreement sharply and decreases the sensitivity of the Cohen's Kappa method. For example, a disagreement on one study in a set of studies with a vast majority of 'excludes' will cause a considerably larger decrease in the Cohen's Kappa coefficient than a disagreement on one study in a set of studies equally balanced between 'includes' and 'excludes'.

After arrival of the hard copies, all 144 studies were screened independently by two team members using the same exclusion criteria. Only in five cases (3%) disagreement existed which were resolved after discussion.

The 66 studies constituting the review map include 15 of the 31 pre-identified key studies. As predicted, the eight background studies were eliminated. In addition, listed key studies were excluded in cases where their focus did not coincide with the review question. For instance, key studies dealing with context-based assessment and with context-based curriculum development were excluded.

Sixty-six studies were keyworded, of which 16 studies (over 20%) were keyworded by at least two team members. An EPPI-Centre member also keyworded six of these. Table 3.7 shows the inter-rater agreement for the keywording process.

Table 3.7: Inter-rater agreement for keywording

	Keywords on core keywording sheet			**Keywords on review-specific keywording sheet**		
	Agreed	Disagreed	Inter-rater agreement (%)	Agreed	Disagreed	Inter-rater agreement (%)
Team member 1 and team member 2 (N=16 studies)	237	8	97	157	11	95
Team member 1 and EPPI member (N=6 studies)	69	18	79	30	32	48
Team member 2 and EPPI member (N=6 studies)	70	17	80	31	34	48

The table shows that agreement of keyword allocation between team members is consistently high for both the general and the review-specific keywords (97% and 95% respectively). The agreement is lower between the respective team members and the EPPI-Centre member, an expert in a field other than science education. Part of this difference may be explained by the fact that the EPPI-Centre member was not involved in the initial moderation of the keywording terms. Some of the differences in coding for the core keywords can be explained by the specific interpretation by the team members of the nature of researcher-manipulated studies. The agreement of less than half (under 50%) for the review-specific keywords seems due to the difference in familiarity with science education learning outcomes, curriculum initiatives and the possible variations in context-based teaching.

4. IN-DEPTH REVIEW: RESULTS

4.1 Selecting the studies for the in-depth review

The mapping described in the previous section revealed a large number of studies in the area of context-based and STS approaches to the teaching of science. It was therefore necessary to identify a subset of studies for an in-depth review. In making decisions about the subset, a number of criteria were applied after detailed discussion amongst the Review Team members. It was decided that the in-depth review would focus on those researcher-manipulated studies using a control group. It was further decided that the in-depth review would focus on studies of pupils following whole courses which used context-based approaches, and also that the studies reported outcomes in relation to both understanding of science and attitudes to science.

Whilst it is desirable to look at the effects of context-based approaches on understanding *or* attitudes (that is, the full complement of 66 papers), time limitations precluded this. Therefore, the application of the above criterion (that is, understanding *and* attitude) has been applied. This yielded five studies for the in-depth review.

4.2 Comparing the studies selected for in-depth review with the studies in the systematic map

Application of the criteria described in section 4.1 yielded five papers for the in-depth review. These are as follows:

- Barber M (2000) A comparison of NEAB and Salters A-level Chemistry: student views and achievements. University of York: Unpublished MA thesis.
- Ramsden JM (1997) How does a context-based approach influence understanding of key chemical ideas at 16+*? International Journal of Science Education* **19:** 697-710.
- Wierstra R (1984) A study on classroom environment and on cognitive and affective outcomes of the PLON curriculum. *Studies in Educational Evaluation* **10:** 272-282.
- Wierstra RF, Wubbels TA (1994) Student perception and appraisal of the learning environment: core concepts in the evaluation of the PLON physics curriculum. *Studies in Educational Evaluation* **20:** 437-455.
- Yager RE, Weld JD (1999) Scope, sequence and co-ordination: the Iowa project, a national reform effort in the USA. *International Journal of Science Education* **21:** 169-194.

Two of these studies were undertaken in the UK (Barber; Ramsden), one in the US (Yager and Weld) and two in the Netherlands (Wierstra; Wierstra and Wubbels).

Four focused on pupils in the 11 to 16 age range (Ramsden; Wierstra; Wierstra and Wubbels; Yager and Weld) and one on pupils in the 17 to 18 age range (Barber).

One has a science focus (Yager and Weld), two have a chemistry focus (Barber; Ramsden) and two have a physics focus (Wierstra; Wierstra and Wubbels).

4.3 Further details of studies included in the in-depth review

Appendix 4.1 provides summary tables of the five studies included in the in-depth review. These tables are based on the information gathered and judgements reached in the data-extraction of the studies. Where a concise summary was made in the papers, the key conclusions in relation to understanding and attitude have been presented in the author’s own words.

4.4 Synthesis of the evidence

4.4.1 Introduction

This section discusses the evidence presented in the five studies.

The studies had the following in common: they were evaluations of researcher-manipulated interventions; they were all controlled trials; they focused on whole courses with a context-based approach; and they reported on both understanding and attitudes. These studies formed a subset of 26 researcher-manipulated studies included in the systematic map, and these 26 studies comprised 40 percent of the total studies.

The interventions being evaluated

The Barber and Ramsden studies focused on two national courses in the UK, *Salters Advanced Chemistry* (for pupils aged 17 and 18) and *Science: the Salters Approach* (for pupils aged 14 to 16). The latter study (Ramsden) elected to concentrate on aspects of chemical understanding within a science course. Both these courses were developed at the University of York in the UK. The Wierstra and the Wierstra and Wubbels studies focused on a national physics development project for pupils aged 15 and 16 in the Netherlands, the *Dutch Physics Curriculum Development Project (Projekt Leerpakketontwikkeling Natuurkunde, [PLON]).* This course was developed at the University of Utrecht in the Netherlands. The Yager and Weld study focuses on a science course for pupils aged 11 to 16 in Iowa in the US, *Scope, Sequence and Continuity (SS&C)*. This course was developed at the University of Iowa. Although the Yager and Weld paper characterises *Scope, Sequence and Continuity* as an STS approach, it was characterised for the purposes of this review as a context-based course because it had many similarities to other courses characterised as context-based, most particularly in that it aimed to cover the range of science ideas normally covered in more conventional science courses.

Overview of the studies

The Barber study was undertaken by a practitioner-researcher in her own institution and, as such, can be characterised as a case study of practice in one setting. Within this institution, pupils were able to elect to follow either a conventional course or a context-based course. The remaining studies were undertaken by university-based researchers who had gathered data from more than one institution. These studies involved gathering comparative data from classes using more conventional courses and classes using context-based courses. Of these, the Yager and Weld study was the largest in scale, involving the pupils of some 180 plus teachers. The Ramsden, Wierstra and Wierstra and Wubbels studies involved much smaller sample sizes of some one, two or three hundred pupils. One further point worth noting is that, with the exception of the Barber study, the author, or at least one of co-authored papers, had been involved in the development of the course which formed the focus of the study. The involvement of those designing the course in its evaluation clearly has the potential to introduce ethical issues to do with potential researcher bias into the design and reporting of evaluation studies. However, there was no evidence in any of the reports of the studies of any elements of researcher bias.

4.4.2 Research design

All the studies involved some form of non-randomised experimental design.

The Barber report contained two linked studies. The first involved using a questionnaire with 120 pupils randomly selected from groups taking a more conventional course and groups taking a context-based course. This was followed up with interviews of a subset of pupils (five following each course), based on what were described as responses of interest in their questionnaires. The second study used a written test administered to a subset of pupils following each course, with the selection being made on the basis that the pupils were in the two class groups taught by the researcher.

The Ramsden study involved two groups of 84 pupils, one following a context-based course and the other more conventional courses. Eight schools were involved in total.

The Wierstra study used class sets in the research design, with 254 pupils in nine classes taking the *PLON* course and 144 pupils in six classes taking a more conventional course.

The Wierstra and Wubbels study also used class sets, with 209 pupils in eight classes taking the PLON course, and 355 pupils in 16 classes taking the more conventional course.

The Yager and Weld study involved a much larger sample than any of the other studies. 5,270 pupils across five school districts were taking the SS&C course, and 1,320 pupils following more conventional courses.

Steps were taken in each of the studies to control some of the variables in constructing samples. In the Barber study, the written test was given to pupils in both groups who were taught by the researcher. In the Ramsden study, the pupils were matched in terms of ability on the basis of their predicted GCSE

grades. In the remaining three studies, data were gathered from groups broadly similar in composition in terms of the schools from which they were drawn.
The research designs point to some of the constraints operating on evaluations of this nature. It is not always possible to allocate pupils or classes randomly to groups which will or will not receive an intervention. Most often, the research design has to be built around existing class sets in schools. The sampling often has an opportunistic dimension, in that schools and classes using the new intervention are identified, and then other schools using more conventional course are identified to create a comparison group of roughly the same size. However, within this, it might be possible to identify subsets of pupils who are more closely matched in terms of particular characteristics and therefore enhance the validity of any findings. The Ramsden study demonstrates one way in which this might be achieved, by gathering data from whole class sets of pupils, but using another measure (predicted grades in a public examination) to construct two groups matched in terms of ability. Other pre-intervention benchmarks which did not exist in the UK at the time of the study (such as Standard Assessment Tasks [SATs]), would also be appropriate here.

4.4.3 Data-collection and analysis instruments used

The Barber study gathered data on understanding in the form of pupils' written answers to 14 questions developed by an external body, the Royal Society of Chemistry, to assess pupils' knowledge and understanding of chemistry. Additional data were also gathered about 'value added': that is, the extent to which the courses appeared to build on the potential demonstrated by pupils at GCSE. Data on attitudes were gathered through a short self-developed questionnaire and semi-structured interview schedule. Piloting of the questionnaire was undertaken to enhance its validity. The questionnaire used a mixture of fixed response items, free response items and agreement/disagreement scales.

The Ramsden study gathered data on understanding via eight diagnostic questions, two on each of four key ideas in chemistry, some of which were drawn from an earlier study. These were supplemented by a small number of free-response questions seeking information on career intentions and the aspects pupils had liked and disliked about their courses. These questions were developed by the researcher. The majority of the diagnostic questions were drawn from previous studies which had involved reliability and validity checks.

In the Wierstra study, data on understanding were gathered using achievement tests based on PLON examinations and conventional physics examinations. Eighteen items from each were included. Data on attitude were gathered using a self-developed attitude inventory developed earlier by the group at Utrecht, and which employed 12 items for students to rate on agreement/disagreement scales. Steps were taken to check on the reliability and validity of the instruments, which drew on instruments developed in earlier studies.

In the Wierstra and Wubbels study, data on understanding were gathered through a self-developed conventional 19-item, multiple-choice physics test covering science knowledge common to PLON and the more conventional courses. Attitude data were gathered using a self-developed questionnaire with agreement/disagreement scales. Correlations were calculated to check on the reliability of measures and a number of measures were employed to check on the validity of the instruments.

In the Yager and Weld study, data were gathered using self-developed instruments previously developed to assess earlier and smaller scale STS interventions developed by the group in Iowa. The instruments involve pupils responding on agreement/disagreement scales to a series of statements in six areas: process, creativity, attitude, world view, concept and application. Data were also gathered on career intention, again using agreement/disagreement scales. Pre- and post-intervention data were collected in the first four of the six areas, and an ‘improvement rate’ formula was used to give a measure of change in the areas of concept and application. Steps were taken to check on the reliability (by test-retest) and validity (through staff panels) of the instruments.

It is clear from the description of these instruments that there is general consensus on the approaches which should be used to gather data, but considerable diversity in the instruments employed. Data on understanding were gathered either through the use of diagnostic questions or, more commonly, through the use of examination and test items drawn from existing banks of items. By far the most common approach to gathering data on attitude is the use of self-developed inventories involving agreement/disagreement scales (Likert-type questionnaires).

The nature of the data gathered means it is not possible to make direct comparisons of studies, nor undertake meta-analyses of data. This raises the question of how feasible it might be to make use of standardised instruments in the evaluation of context-based approaches. Whilst, in principle, this would seem to be a good idea, in practice, a number of problems emerge, particularly when trying to make comparisons of understanding across countries, as surveys such as the Third International Mathematics and Science Survey (TIMMS) in the mid-1990s have demonstrated. Countries differ in their educational frameworks in terms of when pupils start school, the number of years of compulsory schooling, the ages at which pupils sit national tests and examinations, and the curriculum pupils experienced by these points. All these factors mitigate against the validity of using some form of cross-national measure of scientific understanding. It is worth noting, however, that the test items developed for the Organisation for Economic Co-operation and Development / Programme for International Student Assessment), international surveys of scientific literacy – the ‘PISA studies’ – may be of value in gathering cross-national data on the effects of context-based and STS approaches.

The possibility of using some form of standardised measures of understanding within countries would appear to be higher and there might be some merit in compiling a central register of instruments which have been used, particularly if they have been subjected to reliability and validity checks. However, the logistics of this may not be easy to organise.

Attitude instruments would appear to offer fewer problems in terms of use in more than one setting or country. In practice, however, there are very few examples of attitude inventories being used across a range of studies, other than by those working within groups: that is, using instruments they have developed to assess one of their interventions to asses others, as in the case of the instruments used to assess SS&C. On the other hand, researchers tend to develop their own instruments or modify existing instruments in order to come up with a format which they feel will be of particular help in the context of their situation. (This is a reflection of the broader picture in attitudes research, which is characterised by an enormous diversity of research instruments of very variable quality.)

It is worth noting that the Views on Science Technology Society (VOSTS) instrument, originally developed in Canada, has been used in a number of countries to gather data which relates to attitudes. However, it was not used in any of the studies included in the in-depth review. There may be some merit in setting up a central register of instruments which have been used to gather data on attitudes but, again, the logistics of this may not be easy to organise.

Methods used to analyse the data

All the studies were able to draw on the experimental aspects of their design, coupled with the size of their sample groups, to make use of statistical techniques in the analysis of some of the data. The most common were t-tests, calculation of correlation coefficients and analysis of variance. The techniques used in each case were appropriate to the study design. It seems reasonable to conclude from this that researchers working in the area see the use of sample sizes which are sufficient to lend themselves to statistical analysis and the application of such analysis techniques, as an important means of evaluating some aspects of interventions.

4.4.4 The outcomes reported and the conclusions drawn

The Barber study reports that pupils taking the context-based chemistry course (*Salters Advanced Chemistry*) performed noticeably less well than pupils taking more conventional courses on traditional style questions aimed at assessing chemical understanding. However, pupils in both groups achieve a similar range of grades in their end-of-course examinations. (These differ for each group, with the context-based examination using contextualised questions.) The study also reports higher and more sustained levels of interest in their study of chemistry from the pupils following the *Salters Advanced Chemistry* course; more of these pupils electing to go on to study chemistry at university. The value added measures were, in part, a means of checking on the validity of the data on understanding and the interviews provided a means of validating responses on the attitude questionnaire.

The Ramsden study reports little difference in understanding of chemical ideas in pupils following either context-based (Science: The Salters Approach) or more conventional courses, though some ideas were understood poorly by both groups. The pupils following the context-based course expressed higher levels of interest in their science lessons than those following more conventional courses. There was no difference between groups in numbers indicating they wished to study chemistry further. The data on understanding were validated by including two items on each chemical idea in the instrument.

The Wierstra study reports no difference in understanding between pupils following the context-based physics course (PLON) and those following the more traditional course, though PLON pupils performed better in the PLON-style questions on the understanding instrument, and pupils following more conventional courses performed better on more traditional style questions. The study also reported that the inquiry-learning approach associated with PLON resulted in improvements in pupils’ attitudes to science, and also established that attitude correlated with achievement. No details were given of any steps taken to check on the reliability and validity of the data.

The Wierstra and Wubbels study reports no difference in test scores for understanding between pupils taking the context-based physics course (PLON) and those taking more conventional courses. The study also reports affective outcomes as being lower for PLON pupils, contrary to the expectations of the researchers. Statistical tests were applied to the data to check on its reliability and validity.

The Yager and Weld study reports that students taking the context-based course (Scope, Sequence and Continuity, SS&C) show significantly better understanding and more positive attitudes to science than those following more traditional courses, especially in the case of lower ability pupils and female pupils.

Table 4.1 summarises the findings reported in the studies in the in-depth review with respect to pupils taking the context-based courses when compared with pupils following more conventional courses.

The evidence reported in four of the five studies suggests that context-based approaches are at least as effective in developing understanding of scientific ideas as more conventional courses. Four of the five studies report improvements in pupils' affective responses – two to experiences in science lessons, two to science more generally, and the fifth reports a mixture of both positive and negative responses. Thus the evidence as reported in the studies certainly appears to indicate that context-based approaches achieve their aims in terms of providing pupils with positive experiences in their science lessons.

It would be of interest to report the size of the effect of context-based courses. However, only one study (Yager and Weld, 1999) reports some effect sizes. On the other hand, Wierstra (1984) does not present substantial raw data of effect differences but largely concentrates on the relationships between different measures, such as the correlation between affective and cognitive outcomes. Due to the limited scope for a meaningful commentary on size effects, this report includes the reported data on which the authors base their conclusions about the significance of effect differences. Apart from the study with insufficient information (Wierstra, 1988), the reviewers agree that the reported data support the conclusions drawn.

Table 4.1: Summary of the reported effects of context-based courses on understanding and attitude

Study	Effects on understanding	Effects on attitude
Barber (2000)	Poorer performance on test with traditional-style questions *Cognitive test scores (max 75)* *Salters (N 15; mean 37.3)* *Control (N 20; mean 48.2)* similarly close to expected grades in final external examinations *Added Value Perform. Indicator (AVPI)* *Cognitive test scores* *Salters (N 15; AVPI 1.06)* *Control (N 20; AVPI 1.02)*	Higher levels of interest and more sustained interest in chemistry lessons Weighted rank scores, chi-square comparison *Salters (N 60; mean 2.78)* *Control (N 60; mean 2.22)* $\chi^2 = 4.94$, $p>0.05$
Ramsden (1997)	No difference Groups matched on predicted scores *Cognitive test scores (max 20)* *Salters (N 84; mean 9.22; SD 3.82)* *Control (N 84; mean 9.84; SD 3.38)*	Higher levels of recognition of relevance of chemistry lessons to students' lives *Percentage comparison* *Salters 10%; Control 0%*
Wierstra (1984)	No difference on total test Cognitive test scores *PLON (N 254); Control (N 144)* *no actual results provided* but pupils following a context-based course did better on context-based questions *PLON (N 254); Control (N 144)* *PLON higher, p<0.040* and those following a conventional course did better on traditional-style questions *PLON (N 254); Control (N 144) Control higher p<0.001*	More positive attitudes to science *PLON (N 254); Control (N 144)* *no actual results provided* *PLON attitude scores higher, p<0.04.* and attitude scores correlate with achievement *no actual results provided* *correlation coefficient 0.27; p<0.001*
Wierstra and Wubbels (1994)	No difference Cluster design: cognitive test scores (max 19) PLON (N 209; mean 10.17; SD 0.17) Control 1 (N 269; mean 10.36; SD 0.17) Control 2 (N 86; mean 10.93; SD 0.21) Variance between groups p=0.33	Less positive attitudes to science in relation to finding their lessons useful and instructive MANCOVA,Wilks λ 0.97; p=0.007 but lessons seen as more reality-centred and activity-centred e.g. reality-centredness test scores PLON (N 209; mean 2.74; SD 0.58) Control1 (N 269; mean 2.60; SD 0.53) Control2 (N 86; mean 2.46; SD 0.48) Variance between groups p=0.0000 (still significant at cluster level)
Yager and Weld (1999)	Better understanding 'Improvement rate' formula used to adjust cognitive post-test scores, e.g. Y6 (max 100) SS&C (N 1976; mean 72.3; SD 18.3) Control (N 429; mean 49.6; SD 14.8) T=29.1	More positive attitudes to science attitude post-test scores corrected for pre-test as co-variate (ANCOVA), e.g. Y6 (max 100) SS&C (N 1976; mean 60.4; SD 16.8) Control (N 429; mean 50.9; SD 15.9) T=29.4

4.4.5 The quality of the evidence base

The algorithm described in section 2.3.3 (which gives a maximum score of 14) was applied to the studies and Table 4.2 shows the results of the weightings in each category.

Table 4.2: Weightings given to studies in the in-depth review

Study	Category A	Category B	Category C	Category D
Barber (2000)	High (6)	Medium (2)	High (2)	Medium high (10)
Ramsden (1997)	High (6)	Medium (2)	High (2)	Medium high (10)
Wierstra (1984)	Medium (2)	Medium high (4)	High (2)	Medium (6)
Wierstra and Wubbels (1994)	High (6)	Medium high (4)	High (2)	High (12)
Yager and Weld (1999)	High (6)	Medium high (4)	High (2)	High (12)

Unsurprisingly, all the studies obtained high ratings in Category C, which assesses the relevance of the focus of a study for addressing the in-depth review question. In categories A and B, all the studies achieved at least a rating of medium for quality, with the majority having better ratings than this. This indicates that the overall quality of research in the studies is good. It should be noted that the lowest rating for Category A (the Wierstra study) is partly due to the fact that it refers to a separate paper in Dutch (not included in this review) for much of the information on the research methods used. Other factors which lowered the quality ratings were generally associated with limited sample size, limited attempts to check on the reliability and validity of instruments and the data collected, and limitations on the extent to which controls were used in the experimental design.

One feature worthy of note is that the practitioner/researcher study (Barber) achieved a rating at least as good as, or better than, the other studies in most categories.

In making judgements about the studies, it is important to bear in mind what is practicable as well as what is desirable. Evaluations studies interested in the effectiveness of curriculum interventions set out to answer the question 'What works?' and the most rigorous test of 'What works?' is generally seen to be the randomised controlled trial (RCT). The evidence from the studies included in the in-depth review show that this method was not employed for the evaluation of context-based courses. Although a lack of such type of studies does not necessarily indicate that the RCT method is unsuitable, there seem very practical reasons why such trials may not be feasible for evaluating large-scale curriculum interventions. Decisions on participation in such interventions can rarely be made by researchers, but, instead, depend on decisions by policy-makers (ministries) or practitioners (school departments). Researchers then have to work with existing class sets taking given curricula.

More fundamentally, the 'What works?' question, is not as simple as it first appears in the context of educational evaluation. Before it is possible to decide what works, it is necessary to decide what 'working' means. This can be illustrated with reference to the Barber and Wierstra studies. Barber's findings showed that pupils following the context-based course performed significantly less well on standard test items of chemical knowledge and understanding than pupils following more conventional courses. Wierstra showed that pupils following the context-based course did better than pupils following conventional courses on understanding-related assessment items developed for the context-based course, whilst the performance of pupils following conventional courses was superior on traditional-style items.

It seems perfectly reasonable to suggest that, if the aims of an intervention are different, the way in which it should be assessed is also likely to be different. The Barber study established that pupils in both groups achieved similar grades in their A-level examinations. (All A-level courses are assessed by an external group, the Qualifications and Curriculum Authority [QCA] to ensure standards are comparable.) Thus if 'working' means getting similar marks on traditional-style questions, the context-based course clearly does not 'work'. If, however, it means getting similar grades on external examinations judged to be of the same standard, than it does 'work'. Most context-based and STS interventions involve a shift in the intended outcomes, and the old and the new cannot therefore be directly compared. This implies that some aspects of such interventions have to be evaluated against their declared aims, and not by comparisons with another approach.

4.5 In-depth review: quality assurance results

One of the studies in the in-depth review (Barber) is a thesis identified through personal contacts. For this review, no handsearch was done of the Dissertation Abstracts. However, several of the databases included in the search (e.g. ERIC and PsychINFO) list dissertations. The Barber study was not identified through the systematic search of these databases but through personal contacts. The inclusion of a study identified through this route may have introduced some bias to the in-depth review. The sensitivity of the synthesis of the findings to this potential bias is minimal, since the Barber findings are similar to those of the other studies included in the in-depth review, and its weighting is close to the combined weighting of all these studies.

The quality assurance processes for in-depth reviewing described in section 2.3.5 were followed. No areas of significant disagreement remained after moderation of the data-extraction process. Generally, guidelines by collaborators from the EPPI-Centre were followed. Only where explicit criteria for quality judgements differed between experts was consistency across studies established by consultation with the whole Review Team. The algorithm developed to reach the overall weighting for each of the studies worked well in that it discriminated adequately between them. Additionally, all the team members had independently ranked the five papers on the basis of what they felt was the overall quality. Their identical rankings matched those produced by the algorithm.

5. FINDINGS AND IMPLICATIONS

5.1 Summary of principal findings

This review set out to answer the question: **What evidence is there that teaching approaches which emphasise placing science in context and promote links between science, technology and society improve the understanding of science ideas and the attitudes to science of 11 to 18-year-old pupils, and what are the implications of the evidence for initial teacher training courses?**

The in-depth review focused on the question: **What evidence is there from controlled evaluation studies that context-based courses improve understanding of science ideas and the attitudes to science of 11 to 18-year-old pupils, and what are the implications of the evidence for initial teacher training courses?**

The in-depth review has demonstrated that there is strong evidence to support the claim that context-based approaches motivate pupils in their science lessons. The in-depth review has further demonstrated that there is evidence to suggest such approaches also foster more positive attitudes to science more generally. Finally, the in-depth review provides (a) evidence to suggest that context-based approaches do not adversely affect pupils' understanding of scientific ideas and (b) more limited evidence to suggest they can improve understanding.

5.1.1 Identification of studies

The review identified 66 studies which met the inclusion criteria developed for producing the systematic map. These studies all report some form of evaluation of 11- to 18-year-old pupils' understanding in science and/or attitudes to science in relation to courses which teach science through a context-based or science-technology-society (STS) approaches.

5.1.2 The mapping of the included studies

The mapping of the included studies was done with reference to a number of characteristics. These included looking at the location of the report of the study, looking for linked reports to identify clusters of related work, the type of study undertaken, the nature of the intervention, the aspects of science learning the study was addressing, and the outcomes and how these were measured.

A brief summary of the main features of the map is given as follows:

- The map indicated that the majority of reports were identified in the electronic databases ERIC and BEI, and through personal contacts.
- There were clusters of linked papers indicating concentrations of activity in the UK, the US, the Netherlands, Canada and Southern Africa.
- All 66 studies were evaluations, with 26 using some form of control group in the research design.

- Over half the studies (38) were in the discipline of science, with 15 focusing on chemistry, nine on physics, three on biology, and one on both chemistry and physics.
- 46 of the interventions focused on bringing about change in attitudes to science and 44 on understanding of science ideas; 24 focused on both attitudes and understanding.

5.1.3 The nature of the studies selected for the in-depth review

In addition to meeting the criteria for inclusion in the systematic map, the studies selected for the in-depth review all used control groups. These focused on pupils following whole courses which used context-based approaches, and reported outcomes in relation to both understanding of science and attitudes to science.

5.1.4 Synthesis of findings from studies in in-depth review

The synthesis of findings from the in-depth review revealed a number of interesting features. In particular, it suggested that those working in the area saw experimental designs as important in evaluating some aspects of the effectiveness of context-based approaches, and steps were therefore taken to ensure studies contained adequate sample sizes and control of some key variables where feasible. As such, the research was generally of a high or medium high quality, suggesting good levels of confidence in the findings. The synthesis also demonstrated that there was a consensus over the broad strategies which should be employed to assess levels of understanding, but diversity in actual instruments used.

5.2 Strengths and limitations of this systematic review

The review has a number of strengths. Some of these relate to the substantive content, and some to methodological considerations.

5.2.1 The substantive content

Firstly, the strength of the review is the systematic nature of searching, screening, keywording and data-extracting a variety of studies on an area of considerable interest and topicality in science education. The Review Team, being familiar with the area, were pleased that the review helped identify a number of previously unknown quality studies in the area, indicating that the review process had been worthwhile. Secondly, the review has enabled the area to be mapped systematically which, in turn, has enabled key features of the work to be identified, thus creating new knowledge. Thirdly, the in-depth review has demonstrated that research in the area is generally of a quality which means confidence can be placed in its findings.

5.2.2 Methodological considerations

The review has enabled team members to gain valuable perspectives on the features of a research paper which make for clear, concise and accessible reporting of research findings. This has helped Review Team members clarify their views on the reporting of research studies to different audiences. More generally, and in the current climate of heated debate about the nature and purpose of educational research, the Review Team has welcomed the opportunity to consider carefully what counts as 'good' research.

There were a number of limitations to the review. Some of these concern methodological considerations, and others practical matters.

Limitations to coverage: The 'effectiveness-focus' of the review question, in common with those of many other EPPI-style reviews, means that the included studies are solely empirical studies. These clearly have an important role to play in providing evidence to help answer what is often described as the key effectiveness question, 'What works?' However, by only covering reports of empirical studies, the review excludes a very important body of literature which addresses policy matters, theoretical considerations and other issues related to the review focus. Whilst these might not answer the 'what works?' question, and thus directly address the review question, they provide essential context for anyone wanting to familiarise themselves with an area and the key questions people are asking about it. The Science Review Team used an additional coding category when applying the inclusion and exclusion criteria. This involved categorising studies which covered essential background and contextualising information. **The Review Team recommends that the review procedures and the report structure should be modified such that references to such contextual and theoretical papers can be incorporated to provide a more complete picture of the area.**

Characterisation of studies: The options available for characterising studies in Section 10 of the EPPI-Centre Core keywording sheet are, naturally, limited. In the context of the review question, it was inevitable that the studies included in the review were evaluations. Thus a considerable diversity of studies was characterised under only a very limited number of headings. For instance, action research, practitioner research, case studies and various forms of experimentally designed studies were all grouped together as either researcher-manipulated or naturally occurring evaluations. **The Review Team suggests that subsequent reviews make use of the opportunity for including a further review-specific keyword category with options for a refinement of relevant research designs in order to avoid the risk of losing valuable data in the keywording process and to enrich the review map. This category could include, for instance, action research, practitioner research and case study.** Several of the terms used in EPPI reviews for classifying types of studies are unfamiliar to the majority of people undertaking educational research, as judged by the ways in which they characterise their work. (The Review Team did not find a single example of a study characterised by those reporting the work as either a naturally occurring or researcher-manipulated intervention.) It seems only sensible to communicate the findings of reviews in ways which will have some meaning to those reading the reports. **The Review Team suggests that the core EPPI-Centre keywords are modified to include terms more familiar to education researchers.**

Timescale: The experience of the Science Review Team, where team members were often working fulltime on the review, suggests that 11 weeks is too short a period, particularly as there are a number of key rate-determining steps in the process which cannot easily be accelerated, such as the requesting of papers through inter-library loan. Too short a timescale can compromise the quality of the review, by providing insufficient time for careful reflection on the outcomes once the EPPI procedure has been followed. One outcome of the short timescale is that only a very small area of the map formed the basis of the in-depth review. **The Review Team recommends that any further reviews be allocated more time to ensure quality is as high as possible, and reviews have breadth as well as depth.**

The length of the report: The detail required in the report is extensive. Whilst this is appropriate for some audiences, the nature and size of the document (which has been written in accordance with guidelines provided on structure and content) means that it is unlikely to be read by some key potential audiences. **The Review Team supports the efforts made to summarise and disseminate the findings of the review in additional formats suitable for key audiences of ITT practitioners and policy makers.**

5.3 Implications

In considering the implications of the review, it is important to bear in mind that the in-depth review covered only five papers. However, despite this small sample size, there is a high degree of consensus in the findings.

5.3.1 Implications for research

The review has implications for both secondary and primary research.

Secondary research

Highly desirable secondary research in the context of this particular review would involve undertaking in-depth reviews of other areas of the map. These would prove very informative in terms of establishing the breadth of the evidence base on which claims are made for context-based and STS approaches in science teaching. In particular, areas of the map which would benefit from in-depth review are researcher-manipulated interventions which focused on whole STS courses, and attitudes alone or understanding alone. Additionally, many of the papers in the map make claims for beneficial effects on disadvantaged pupils, pupils in specific ability ranges and pupils of different gender. These areas would also be candidates for in-depth review. However, the Review Team does not feel that the area of skills is worth pursuing in detail. Despite claims being made in many of the studies to develop ‘skills’ of one form or another, the term was used to describe a very wide range of attributes, many of which appeared to be very difficult to measure with confidence.

Given that the review used a tight definition of researcher-manipulated evaluations as those including a control group, it would also be important and interesting to conduct an in-depth review of some of the uncontrolled evaluation studies in order to assess the nature and quality of the evidence they present.

Primary research

As well as considering what is in the map, it is also important to look for what is not there, as this points to areas where primary research may be of particular value. The terms 'context-based approaches' and 'STS approaches' cover a diversity of content and incorporate a wide range of activities, some of which are not traditionally associated with science teaching. Reviews focusing on the effects of particular types of teaching activity would be informative. In this context, it is worth noting that the first review of the EPPI Review Group for Science is focusing on the effects of small-group discussions. A second area would be to look in detail at the effects of culture on understanding and attitude when learning through a contexts-based or STS approach.

Given the problems identified in comparing understanding discussed in section 4.4.3, any moves towards developing standardised instruments for the purposes of making comparisons are likely to be more fruitful if directed at instruments to assess attitudes. Such instruments would also be helpful in evaluating scientific literacy courses. These will clearly introduce different ideas in science lessons, making comparisons on the basis of understanding of ideas problematic. However, one aim of these courses different is to foster more positive attitudes in pupils, and here comparisons can clearly be made with other types of course.

5.3.2 Implications for policy and practice in initial teacher training

The review has a number of implications for policy and practice in initial teacher training in England and Wales. The use of context-based approaches (to use the term more commonly employed in these countries) has been advocated for a number of years in conjunction with the associated 'active learning' strategies as a means of motivating pupils in lessons. The review indicates that the evidence base for such a recommendation is sound. It is therefore essential that those on initial teacher training courses should be aware of this, as should those who produce materials used on such courses. These include bodies such as the DfES, who are responsible for introducing the *Key Stage 3 Strategy*, the QCA, who have produced an extensive and detailed *Scheme of Work for Key Stage 3*, and the TTA, who specify standards for newly qualified teachers. This is particularly important given the current move towards developing school science courses which emphasise scientific literacy and the considerable areas of overlap in terms of both content and approach between such courses and context-based/STS courses.

Another key audience for the findings of the review are the Office for Standards in Education (Ofsted), as they make judgements about the quality of lessons they observe. There is increasing anecdotal evidence that many science lessons now begin with pupils copying down into their books the intended learning outcomes of the lesson from the board or an overhead projector transparency. Whilst it is certainly important that both initial teacher training students and pupils are clear about what they hope pupils will learn in their lessons, this is a necessary but not sufficient condition for high quality science teaching. It is clear from the review that context-based approaches provide an effective way to interest and motivate pupils in their science lessons.

6. REFERENCES

6.1 Studies included in the systematic map

**Indicates studies included in the in-depth review.*

Alsop S, Watts M, Hanson J (1998) Pupils' perceptions of radiation and radioactivity: the wary meet the unsavoury. *School Science Review* **79:** 75-79.

Banks P (1997) *Students' Understanding of Chemical Equilibrium*. University of York, UK: Unpublished MA thesis.

*Barber M (2000) A comparison of NEAB and Salters A-Level chemistry: students' views and achievements. University of York, UK: Unpublished MA thesis.

Barker V, Millar R (1996) *Differences Between Salters' and Traditional A-level Chemistry Students' Understanding of Basic Chemical Ideas*. York, UK: University of York.

Barker V, Millar R (1999) Students' reasoning about chemical reactions: what changes occur during a context-based post-16 chemistry course? *International Journal of Science Education* **21:** 645-665.

Barker V, Millar R (2000) Students' reasoning about basic chemical thermodynamics and chemical bonding: what changes occur during a context-based post-16 chemistry course? *International Journal of Science Education* **22:** 1171-1200.

Ben-Zvi R (1999) Non-science oriented students and the second law of thermodynamics. *International Journal of Science Education* **21:** 1251-1267.

Campbell B, Hogarth S, Lubben F (2000a) Contextualising the physics curriculum: learners' perceptions of interest and helpfulness. Paper presented at the Annual Meeting of the British Educational Research Association. Cardiff, UK: September 7-9.

Campbell B, Lubben F, Dlamini Z (2000b) Learning science through contexts: helping pupils make sense of everyday situations. *International Journal of Science Education* **22:** 239-252.

Christie DJ, Nelson L (1988) Student reactions to nuclear education. *Bulletin of the Atomic Scientists* **44:** 22-23.

Dahncke H, Behrendt H, Reiska P (2001) A comparison of STS-teaching and traditional physics lessons on the correlation of physics knowledge and taking action. In: Dahncke H, Behrendt H (eds) *Research in Science Education - Past, Present and Future*. Netherlands: Kluwer Academic Publishers, pages 77-82.

Dlamini B, Lubben F, Campbell B (1996) Liked and disliked learning activities: responses of Swazi students to science materials with a technological approach. *Research in Science and Technological Education* **14:** 221-236.

Ebenezer JV, Zoller U (1993) Grade 10 students' perceptions of and attitudes toward science teaching and school science. *Journal of Research in Science Teaching* **30:** 175-186.

Eijkelhof H, Lijnse P (1988) The role of research and development to improve STS education: experiences from the PLON project. *International Journal of Science Education* **10:** 464-474.

Ferreira M (2001) Building communities through role models, mentors, and hands-on-science. *School Community Journal* **11:** 27-37.

Fowler C, Sinclair A (1992) Forecasting the future: a student vision quest. *Science Teacher* **59:** 50-53.

Gallagher SA, Stepien WJ, Rosenthal H (1992) The effects of problem-based learning on problem-solving. *Gifted Child Quarterly* **36:** 195-200.

Greenlee S, Lambert L (1996) *The Science of Living Spaces: Women in the Environment of the 21st Century. An Evaluation and Guide Book.* Virginia, USA: Christopher Newport University.

Helms JV (1998) Science and/in the community: context and goals in practical work. *International Journal of Science Education* **20:** 643-653.

Hughes G (2000a) Marginalization of socioscientific material in science-technology-society science curricula: some implications for gender inclusivity and curriculum reform. *Journal of Research in Science Teaching* **37:** 426-440.

Hughes G (2000b) Salters' curriculum projects and gender inclusivity in science. *School Science Review* **81:** 85-89.

Huppert J, Simchoni D, Lazarowitz R (1992) Human health and science: a model for an STS high school biology course. *American Biology Teacher* **54:** 395-400.

Key M-B (1998) *Students' Perceptions of Chemical Industry; Influences of Course Syllabi, Teachers, Firsthand Experience*. York, UK: University of York.

Kortland J (1997) Garbage: dumping, burning and reusing/recycling: students' perception of the waste issue. *International Journal of Science Education* **19:** 65-77.

Kortland J (2001) A problem-posing approach to teaching for scientific literacy: the case of decision-making about packaging waste. In: De Jong O, Savelsbergh JR, Alblas A (eds) *Teaching for Scientific Literacy*. Utrecht, Netherlands: CDB Press, pages 87-98.

Lenton GM (1990) SATIS 16-19: a preliminary review of school trials. *School Science Review* **71:** 135-140.

Lenton GM (1991) A review of SATIS 16-19 national trials. *School Science Review* **73:** 7-18.

Licht P (1990) Teaching physics and context in the Netherlands. In: Krabbedam H, De Vries M (eds) *Goals and Methods in Science, Mathematics and*

Technology Education. Enschede, Netherlands: National Institute for Curriculum Development, pages 65-78.

Lubben F, Campbell B, Dlamini B (1996) Contextualizing science teaching in Swaziland: some student reactions. *International Journal of Science Education* **18:** 311-320.

Lubben F, Campbell B, Dlamini B (1997) Achievement of Swazi students learning science through everyday technology. *Journal of the Southern African Association for Research in Mathematics and Science Education* **1:** 26-40.

Monhardt RM, Monhardt L (1998) Facilitating science literacy in a rural school. *Bulletin of Science, Technology and Society* **18:** 47-53.

Nentwig P, Parchmann I, Demuth R, Graesel C, Ralle B (2002) Chemie im kontext - from situated learning in relevant contexts to a systematic development of basic chemical concepts. Paper presented at the Symposium on Context-Based Science Curricula. Kiel, Germany: October 10-13.

Pedersen JE (1992) The effects of a cooperative controversy, presented as an STS issue, on achievement and anxiety in secondary science. *School Science and Mathematics* **92:** 374-380.

Phillips LM, Norris SP (1999) Interpreting popular reports of science: what happens when the reader's world meets the world on paper? *International Journal of Science Education* **21:** 317-327.

Pilling G, Holman J, Waddington D (2001) The Salters' experience. *Education in Chemistry* **38:** 131-133.

Ramsden JM (1992) If it's enjoyable, is it science? *School Science Review* **73:** 65-71.

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

Rannikmae M (2001) Effectiveness of teacher-developed scientific and technological literacy materials. In: De Jong O, Savelsbergh JR, Alblas A (eds) *Teaching for Scientific Literacy*. Utrecht, Netherlands: CDB Press, pages 71-86.

Reid N, Skryabina E (2002) Attitudes towards physics. *Research in Science and Technological Education* **20:** 67-81.

Roth WM (2001) Learning science through technological design. *Journal of Research in Science Teaching* **38:** 768-790.

Rubba PA (1990) STS education in action: what researchers say to teachers. *Social Education* **54:** 201-203.

Rubba PA, McGuyer M, Wahlund TM (1991) The effects of infusing STS vignettes into the genetics unit of biology on learner outcomes in STS and genetics: a report of two investigations. *Journal of Research in Science Teaching* **28:** 537-552.

Smith LA, Bitner BL (1993) Comparison of formal operations: students enrolled in ChemCom versus a traditional chemistry course. Paper presented at the Annual Meeting of the National Science Teachers Association. Kansas City, MO, USA: April 1-4.

Smith G, Matthews P (2000) Science, technology and society in transition year: a pilot study. *Irish Educational Studies* **19:** 107-119.

Solomon J (1988) Science technology and society courses: tools for thinking about social issues. *International Journal of Science Education* **10:** 379-387.

Solomon J (1989) Discussing nuclear power. *Physics Education* **24:** 344-347.

Solomon J (1992) The classroom discussion of science-based social issues presented on television: knowledge, attitudes and values. *International Journal of Science Education* **14:** 431-444.

Stevens LP (2001) 'South Park' and society: instructional and curricular implications of popular culture in the classroom. *Journal of Adolescent and Adult Literacy* **44:** 548-555.

Sutman F, Bruce M (1992) Chemistry in the community - Chemcom. *Journal of Chemical Education* **69:** 564-567.

Truex RT (1987) *Township of Ocean School District Contemporary Science. Program description.* New Jersey, USA: Ocean Township Board of Education.

Tsai C-C (1999) The progression toward constructivist epistemological views of science: a case study of the STS instruction of Taiwanese high school female students. *International Journal of Science Education* **21:** 1201-1222.

Tsai C-C (2000) The effects of STS-oriented instructions on female tenth graders' cognitive structure outcomes and the role of student scientific epistemological beliefs. *International Journal of Science Education* **22:** 1099-1115.

Van Weelie D (2001) Contextualising biodiversity. In: De Jong O, Savelsbergh JR, Alblas A (eds) *Teaching for scientific literacy*. Utrecht, Netherlands: CDB Press, pages 99-116.

Walker D (1990) The evaluation of SATIS. *School Science Review* **72:** 31-39.

*Wierstra RFA (1984) A study on classroom environment and on cognitive and affective outcomes of the PLON-curriculum. *Studies in Educational Evaluation* **10:** 273-282.

*Wierstra RFA, Wubbels T (1994) Student perception and appraisal of the learning environment: core concepts in the evaluation of the PLON physics curriculum. *Studies in Educational Evaluation* **20:** 437-455.

Wiesenmayer RL, Rubba PA (1999) The effects of STS issue investigation and action instruction versus traditional life science instruction on seventh grade students' citizenship behaviors. *Journal of Science Education and Technology* **8:** 137-144.

Wiley DA (1991) Observations on a science-technology-society course for non-science students. *Perceptual and Motor Skills* **72:** 845-846.

Williams RA, Lisowski M, Niemietz K (1990) *The Illinois Rivers Project.* Illinois, USA: Southern Illinois University of Edwardsville.

Winther AA, Volk TL (1994) Comparing achievement of inner-city high school students in traditional versus STS-based chemistry courses. *Journal of Chemical Education* **71:** 501-505.

Yager RE, Blunck SM, Binadja A, McCormick RW, Penick JE (1988) *Assessing Impact of S/T/S Instruction in 4-9 Science in Five Domains.* Iowa, USA: University of Iowa.

Yager RE, Tamir P (1993) STS approach: reasons, intentions, accomplishments, and outcomes. *Science and Education* **77:** 637-658.

*Yager RE, Weld JD (1999) Scope, sequence and co-ordination: The Iowa Project, a national reform effort in the USA. *International Journal of Science Education* **21:** 169-194.

Zoller U, Donn S, Wild R, Beckett P (1991) Students' versus their teachers' beliefs and positions on science/technology/society-oriented issues. *International Journal of Science Education* **13:** 25-36.

Zoller U, Ebenezer JV, Morley K, Paras S, Sandberg V, West C, Wolthers T, Tan SH (1988) *Goals' Attainment in Science-Technology-Society (S/T/S) Education: Expectations and Reality. A probe into the case of British Columbia.* British Columbia, Canada: Science Education Research.

Zoller U, Ebenezer JV, Morley K, Paras S, Sandberg V, West C, Wolthers T, Tan SH (1990) Goal attainment in science-technology-society (S/T/S) education and reality: the case of British Columbia. *Science Education* **74:** 19-36.

6.2 Other references used in the text of the report

Aikenhead G (1994) What is STS teaching? In: Solomon J, Aikenhead G (eds) *STS Education: International Perspectives On Reform.* New York: Teachers College Press.

American Chemical Society (ACS) (1988) *ChemCom: Chemistry in the Community*. Dubuque, IA, USA: Kendall-Hunt.

Barker V, Millar R (2000) Students' reasoning about basic chemical thermodynamics and chemical bonding: what changes occur during a context-based post-16 chemistry course? *International Journal of Science Education* **22:** 1171-1200.

Bennett J (2003) *Teaching and Learning Science: A Guide to Recent Research and its Applications.* London: Continuum.

Bennett J, Holman J (2003) Context-based approaches to the teaching of chemistry: what are they and what are their effects? In: Gilbert J (ed) *Chemical*

Education: Towards Research-Based Practice. Dordrecht, Netherlands: Kluwer Academic Publishers.

Borgford C (1995) The Salters Science materials: a study of teachers' use and areas of focus. Unpublished DPhil thesis: University of York.

Burton W, Holman J, Pilling G, Waddington D (1994) *Salters Advanced Chemistry*. Oxford: Heinemann.

Bybee R (ed) (1985) *Science-Technology-Society. 1985 NSTA Yearbook.* Washington DC: National Science Teachers Association.

Campbell R, Lazonby J, Millar R, Nicolson P, Ramsden J, Waddington D (1994) Science: the Salters approach; a case study of the process of large scale curriculum development. *Science Education* **78:** 415-447.

Davies F, Greene T (1984) *Reading for Learning in the Sciences.* Edinburgh: Oliver Boyd.

Dutch Physics Curriculum Development Project (Projekt Leerpakketontwikkeling Natuurkunde [PLON]) (1988) Utrecht, Netherlands: Rijksuniversteit Utrecht, Vakgroep Natuurkunde-Didactiek.

Eijkelhof H, Kortland K (1988) Broadening the aims of physics education. In: Fensham P (ed) *Development and Dilemmas in Science Education.* New York: Falmer Press.

EPPI-Centre (2002a) *EPPI-Centre Core Keywording Sheet* (version 0.9.6). London: EPPI-Centre, Social Science Research Unit.

EPPI-Centre (2002b) *Core Keywording Strategy: Data Collection for a Register of Educational Research* (version 0.9.6). London: EPPI-Centre, Social Science Research Unit.

EPPI-Centre (2002c) EPPI-Reviewer (version 2.0) (Web edition, EPPI-Centre software). London: EPPI-Centre, Social Science Research Unit.

Fensham P (1988) Approaches to the teaching of STS in science education. *International Journal of Science Education* **10:** 346-356.

Hofstein A, Aikenhead G, Riquarts K (1988) Discussions over STS at the Fourth IOSTE Symposium. *International Journal of Science Education* **10:** 357-366.

House of Commons (2002) *Science Education from 14-19. Third Report of the Science and Technology Committee.* London: The Stationery Office.

Hunt A, Millar R (eds) (2000) *AS Science for Public Understanding*. Oxford: Heinemann Educational.

Kyriacou C (1998) *Essential Teaching Skills.* 2nd edition. Cheltenham: Stanley Thornes.

Lewis J (1981) *Science in Society*. London: Heinemann Educational Books.

Millar R, Osborne J (eds) (1998*) Beyond 2000: Science Education for the Future*. London: King's College.

Newton P, Driver R, Osborne J (1999) The place of argumentation in the pedagogy of school science. *International Journal of Science Education* **21:** 553-576.

Nuffield Foundation (1980) *Nuffield Science 13-16.* York: Longman.

Osborne J, Erduran S, Simon S, Monk M (2001) Enhancing the quality of argument in school science. *School Science Review* **82:** 63-70.

Ramsden J (1992) If it's enjoyable, is it science? *School Science Review* **73:** 65-71.

Supported Learning in Physics Project (SLIP) (1997) *Eight Units for Advanced Level.* Buckingham: Open University Press.

Tal R, Dori Y, Keiny S, Zoller U (2001) Assessing conceptual change of teachers involved in STES education and curriculum development – the STEMS project approach. *International Journal of Science Education* **23:** 247-262.

University of York Science Education Group (UYSEG) (1990-1992) Science: the Salters Approach: 22 Unit Guides for Key Stage 4. York/Oxford: UYSEG/Heinemann Educational.

Wray D, Lewis M (1997) *Extending Literacy: Children Reading and Writing Non-Fiction.* London: Routledge.

Yager R, Casteel J (1968) The University of Iowa science and culture project. *School Science and Mathematics* **67:** 412-416.

APPENDIX 2.1: Inclusion and exclusion criteria

TTA systematic review: inclusion/exclusion criteria

Review area: the effect on students of teaching approaches which emphasise placing science in context or promoting links between science, technology and society (STS).

Exclusion criteria: to be included, a study must NOT fall into any one of the following categories:

EXCLUSION ON SCOPE

1. **Not reporting on learning/teaching of science**
   - definition of science: one or several of the school subjects integrated/general science, science, biology, chemistry physics or earth science.

2. **Not about context-based or STS learning/teaching**
   - definition of context-based: using learners' environment and experiences as starting points for teaching);
   - definition of STS: intending to promote links between science and technology or between science and society.

3. **Not about the effect on students' understanding or attitude**

4. **Not about learners aged 11 to 18, or main focus not on learners aged 11 to 18**

EXCLUSION ON STUDY TYPE

(a) Editorials, commentaries, book reviews
(b) Policy documents
(c) Resources
(d) Bibliography
(e) Theoretical (non-empirical) paper
(f) Methodology paper
(g) Review papers

EXCLUSION ON SETTING IN WHICH STUDY WAS CARRIED OUT

6. **Not published in English**

7. **Not published in the period 1980 — 2002**

APPENDIX 2.2: Search strategy for electronic databases

Subject
Teaching approaches which emphasise placing science in context and promote links between science, technology and society.

Population
Pupils aged 11 to 18

Limits
English language
1980 to date.

1. Databases

1a. ERIC

ERIC was searched on 10/02/03 using the BIDS Ovid interface. The database was searched for the period 1980 – 2002 and 1,928 records were retrieved.

1. *(context-based or context-led or application-based or application-led)*
2. (contextualis$ or contextualiz$ or technological approach)
3. (Salters or SATIS or PLON or ChemCom or Science in context)
4. *(science-technology-society or (science and society))*
5. 1 or 2 or 3 or 4
6. 5 and (science or biology or chemistry or physics or earth science)
7. 6 and (course or courses or program$ or project$ or material$ or curricul$ or intervention$)
8. limit 7 to (English language and (elementary secondary education or elementary education or intermediate grades or secondary education or middle schools or junior high schools or high schools or high school equivalency programs or postsecondary education or two year colleges) and (books or 'collected works (general and serials)' or conference proceedings or dissertations or 'evaluative or feasibility reports' or general reports or journal articles or project descriptions or 'research or technical reports' or 'speeches or conference papers') and yr=1980-2002)

1b British Education Index (BEI)

BEI was searched on 11/02/03 using the BIDS Ovid interface. The database was searched for the period 1980 – 2002 and 176 records were retrieved.

1. (context-based or context-led or application-based or application-led)
2. (contextualis$ or contextualiz$ or technological approach)
3. (Salters or SATIS or PLON or ChemCom or Science in context)
4. (science-technology-society or (science and society))
5. (out-of-school experience or everyday or authentic)
6. 1 or 2 or 3 or 4 or 5 or 6 and (science or biology or chemistry or physics or earth science)
7. limit 7 to English language

1c PsycINFO

PsycINFO was searched on 12/02/03 using the WINSPIRS interface. The database was searched for the period 1980 – 2002 and 33 records were retrieved.

1. context-based or context-led or application-based or application-led
2. contextualis* or contextualiz* or technological approach
3. Salters or SATIS or PLON or ChemCom
4. science-technology-society or science-technology-society or sciencetechnologysociety or science-through-technology or socioscientific or socio-scientific
5. #1 or #2 or #3 or #4
6. #5 and (education* or school* or college or student* or pupil* or learner*) and (science or biology or chemistry or physics or earth science)
7. #5 and (education* or school* or college or student* or pupil* or learner*) and (science or biology or chemistry or physics or earth science)
8. #7 and (course or courses or program* or project* or curricul* or material* or intervention*)
9. #8 and (LA:PY = English) and ((PT:PY = case-study) or (PT:PY = clinical-trial) or (PT:PY = collected-works) or (PT:PY = conference-proceedings-symposia) or (PT:PY = empirical-study) or (PT:PY = experimental-replication) or (PT:PY = followup-study) or (PT:PY = interview) or (PT:PY = journal-abstract) or (PT:PY = literature-review-research-review) or (PT:PY = longitudinal-study) or (PT:PY = meta-analysis) or (PT:PY = program-evaluation) or (PT:PY = prospective-study) or (PT:PY = retrospective-study) or (PT:PY = treatment-outcome-study)) and (PY:PY = 1980 – 2002)

1d. Social Science Citation Index

The Social Science Citation Index was searched on 12/02/03 using the Web of Science interface. The database was searched for the period 1992 – 2003 and 210 records were retrieved. It was not possible to download search strings with the records, but these largely replicated those used for ERIC and BEI.

2. Handsearches

The journal *Studies in Science Education* was handsearched as it contains major review articles and was seen as a potential key source of studies

APPENDIX 2.3: EPPI-Centre keywording sheet

Note: Please refer to the EPPI-Centre Core Keywording Strategy (version 0.9.6) for guidance on how to apply keywording.

1. Identification of report
Citation
Contact
Handsearch
Unknown
Electronic database *(Please specify.)*
……………………………………………

2. Status
Published
In press
Unpublished

3. Linked reports
Is this report linked to one or more other reports in such a way that they also report the same study?

Not linked
Linked *(Please provide bibliographical details and/or unique identifier.)*
…………………………………………………
…………………………………………………
…………………………………………………

4. Language *(Please specify.)*

………………………………………..

5. In which country/countries was the study carried out? *(Please specify.)*
……………………………………………
……………………………………………

6. What is/are the topic focus/foci of the study?
Assessment
Classroom management
Curriculum*
Equal opportunities
Methodology
Organisation and management
Policy
Teacher careers
Teaching and learning
Other *(please specify)* ………………………

***6a Curriculum**
Art
Business Studies
Citizenship
Cross-curricular
Design & Technology
Environment
General
Geography
Hidden
History
ICT
Literacy – first language
Literacy further languages
Literature
Maths
Music
PSE
Physical Education
Religious Education
Science
Vocational
Other *(Please specify.)*……………………..

7. Programme name *(Please specify.)*

…………………………………………………

8. What is/are the population focus/foci of the study?
Learners*
Senior management
Teaching staff
Non-teaching staff
Other education practitioners
Government
Local education authority officers
Parents
Governors
Other *(Please specify.)*…………………………………

***8a Age of learners (years)**
0-4
5-10
11-16
17-20
21 and over

***8b. Sex of learners**
Female only
Male only
Mixed sex

9. What is/are the educational setting(s) of the study?
Community centre
Correctional institution
Government department
Higher education institution
Home
Independent school
Local education authority
Nursery school
Post-compulsory education institution
Primary school
Pupil referral unit
Residential school
Secondary school
Special needs school
Workplace
Other educational setting *(Please specify.)*…………..

10a. Which type(s) of study does this report describe?

A. Description
B. Exploration of relationships
C. Evaluation
   a. Naturally occurring
   b. Researcher-manipulated*
D. Development of methodology
E. Review
   a. Systematic review
   b. Other review

see 10b.*

10b. To assist with the development of a trials register, please state if a researcher-manipulated evaluation is one of the following:
Controlled trial (non-randomised)
Randomised controlled trial (RCT)

Please state here if keywords have not been applied from any particular category (1-10) and the reason why (e.g. no information provided in the text):

…………………………………………
…………………………………………
…………………………………………
…………………………………………
…………………………………………
…………………………………………

PTO to apply review-specific keywords (if applicable)

Review-specific keywording sheet

For each item tick any number of keywords **11. What discipline?** a. (integrated) Science b. Biology c. Chemistry d. Physics e. Earth Science f. Environmental Science **12. What types of learners are involved?** a. mixed ability b. lower ability / slow learners c. middle ability d. upper ability / gifted e. disaffected f. non-science	**13. What is the nature of the intervention?** a. full context-based course b. context-based unit/module c. full STS course d. STS enrichment topics **14. What aspect of science learning does the intervention focus on improving?** a. Attitudes *(e.g. attitude to science, attitude to school science, motivation to learn, interest in science activities, social/group collaboration, career intentions)* b. Understanding of science *(e.g. science concepts, the nature of science, scientific methods)* c. Skills *(e.g. investigative skills, manipulative skills, communication skills, problem solving skills, decision making skills)*	**15. How are outcomes measured?** a. test results b. external examination results c. written reports/ open questionnaires d. concept webs e. (dis)agreement scores (incl VOSTS) f. self-reports *(e.g. diaries, interviews)* g. recorded group discussions h. presentations i. observed behaviour **16. Outcomes report on effectiveness in terms of:** a. understanding of science b. attitudes to science c. skills **for learners of different:** d. ability (lower/middle/higher) e. attitude (incl. disaffected) f. science specialisation (incl. non-sc) g. gender h. educational level

Other remarks:

APPENDIX 4.1: Summary tables of studies included in the in-depth review

<table>
<tr><td colspan="2">Barber, M (2000) A comparison of NEAB and Salters A-level Chemistry: student views and achievements.
University of York: Unpublished MA thesis.</td></tr>
<tr><td>Country</td><td>UK: England</td></tr>
<tr><td>Nature of intervention</td><td>Full context-based course</td></tr>
<tr><td>Name of course</td><td>Salters Advanced Chemistry</td></tr>
<tr><td>Age range of learners</td><td>17 to 18</td></tr>
<tr><td>Science discipline</td><td>Chemistry</td></tr>
<tr><td>Type of study</td><td>Researcher-manipulated evaluation, non-randomised controlled trial</td></tr>
<tr><td>Aims of study</td><td>To compare the views of pupils following a conventional Advanced level (A-level) chemistry course (NEAB [Northern Examinations and Assessment Board] Chemistry) and pupils following a context-based course (SAC [Salters Advanced Chemistry]) on a range of different aspects concerning their study of chemistry, and to monitor their achievements and destinations at the end of the course.</td></tr>
<tr><td>Details of sample</td><td>This thesis reports four studies in total, of which two have been included in the review:
1. Study of pupils’ views on their A-level (Advanced-level) course
2. Study of performance on selected test questions.
Study 1 comprises two studies:
1a. Questionnaire study
1b. Follow-up interviews
Sample sizes are as follows:
Study 1a: 120 pupils, 60 from each course selected at random from four class sets following each course
Study 1b: 10 pupils from the 20 who completed the questionnaire, five from each course
Study 2: 35 pupils</td></tr>
<tr><td>Methods used to collect data</td><td>Study 1a: Questionnaire (agree/disagree scales and free-response questions, self-developed and piloted)
Study 1b: Interview (semi-structured format; self-developed)
Study 2: Written test based on test devised by Royal Society of Chemistry, grades on external examination, measures of ‘value added’</td></tr>
<tr><td>Data collected</td><td>Study 1a: Responses to questionnaires
Study 1b: Interview data
Study 2: Responses to 14 items on written test, grades on external examination, measures of ‘value added’</td></tr>
<tr><td>Methods used to analyse data</td><td>Study 1a: Categorisation of responses, some limited statistical analysis (t-test)
Study 1b: Coding scheme to categorise interview responses
Study 2: Numerical marks obtained for questions on written test; comparison using correlation coefficients of marks with GCSE (General Certificate of Secondary Education) and A-level grades on external examination; comparison of GCSE and A-level grades to provide measure of ‘value added’</td></tr>
<tr><td>Summary of results</td><td>Study 1a
• NEAB pupils tended to select the course as a ‘safe option’ because of career intentions; SAC pupils selected the course for interest in the subject and its approach.
• SAC pupils reported finding chemistry harder than NEAB pupils, despite having better GCSE grades, and were less positive about their</td></tr>
</table>

	abilities to cope. Study 1b • This confirmed the questionnaire findings and, additionally: • NEAB pupils had lost some of their enthusiasm by the end; SAC pupils had retained it and felt the course was enjoyable, challenging and helped them learn useful study skills. Study 2 • SAC pupils got lower scores on the RSC-based test than NEAB pupils. • SAC pupils got very slightly better A-level grades than NEAB pupils. • The RSC test has a better 'fit' in terms of style of questions with NEAB assessment than SAC assessment. • 'Value added' is the same for both groups, indicating that the standard of A-level exam is similar for both courses. Reported elsewhere in the thesis: • A higher proportion of SAC pupils go on to study chemistry at university than NEAB pupils.
Conclusions in relation to understanding	'NEAB students appear to cope better with questions designed to assess basic understanding than do Salters students of similar ability.' (p 89) The main criticism of this type of test is that it only assesses a narrow range of skills, namely recall and application of knowledge and understanding of concepts. However, as Barker and Millar point out, the Salters course tests and develops a range of other skills…' (pp 67-68) 'Students taking either course have the same chance of achieving a particular grade.' (p 75) 'There is little difference between NEAB and SAC A-level results in terms of added value performance indicators.' (p 89)
Conclusions in relation to attitudes	'The main reason why NEAB students choose A-level chemistry is because they need it for their future career, whereas Salters students tend to choose chemistry because they are interested in it. The main reason why NEAB students choose the syllabus is because it follows on from their GCSE course, whereas for Salters students it is because they are interested in seeing how chemistry relates to the outside world. Salters students seem to find it harder to adapt to the A-level course, and show an initial lack of confidence in their ability to be successful. However, once settled, they appear to maintain their interest and enjoyment of the subject longer than the NEAB students'. (p 89)
Weight of evidence A: Trustworthiness of study findings in relation to answering study question	High
Weight of evidence B: Appropriateness of research design and analysis for addressing review questions	Medium
Weight of evidence C: Relevance of focus of study for addressing review questions	High
Weight of evidence D: Overall weight of evidence	Medium high

Ramsden J M (1997) How does a context-based approach influence understanding of key chemical ideas at 16+? *International Journal of Science Education* 19: 697-710.	
Country	UK: England
Nature of intervention	Full context-based course
Name of course	Science: The Salters Approach
Age range of learners	15 to 16
Science discipline	Chemistry (The study selected one discipline area within a science course.)
Type of study	Experiment: Researcher-manipulated evaluation, non-randomised controlled trial
Aims of study	To explore the effects of a context-based approach on pupils' understanding of some key chemical ideas at 16+. To explore pupils' views on enjoyment of lessons and career intentions.
Details of sample	168 pupils in eight schools, four using context-based course and four using conventional courses; 84 middle and upper ability pupils in each of the control and experimental groups, matched for ability on the basis of predicted GCSE grades.
Methods used to collect data	• Written questionnaire containing eight diagnostic questions • Free-response questions exploring enjoyment of science and career intentions • Data on projected pupil grades in forthcoming public examinations (General Certificate of Secondary Education, GCSE)
Data collected	• Responses to eight diagnostic questions plus free response items • Projected GCSE grades
Methods used to analyse data	• Numerical rating of responses to diagnostic questions • Statistical analysis (standard deviation, standard error of the mean, critical ratio) • Coding of responses to free response questions
Summary of results	• No significant differences were established in performance on diagnostic questions. • Some key ideas were poorly understood as evident from written explanations, irrespective of approach. • 20 percent of pupils indicated they wanted to study chemistry beyond the compulsory period, equally divided between each approach. • The majority of those not pursuing their study further felt chemistry was not relevant for their career intentions. • Pupils following the context-based course reported higher levels of interest in their science lessons than those following the more conventional course.
Conclusions in relation to understanding	'…this study suggests that there is little difference in promoting understanding of key chemical ideas between a context-based approach and a more traditional approach'. (p 710)
Conclusions in relation to attitudes	'In terms of pupils' more general responses to science, their comment indicate that a context-based approach appears to be enjoyed by those following such a course and makes them more interested in what they are studying. None-the-less, it remains a concern that further study of chemistry beyond the compulsory period is not seen as a worthwhile route to take by the majority of pupils, irrespective of the course they follow, where choices are clearly made on the basis of career aspirations'. (p 710)
Weight of evidence A: Trustworthiness of study	High

findings in relation to answering study question	
Weight of evidence B: Appropriateness of research design and analysis for addressing review questions	Medium
Weight of evidence C: Relevance of focus of study for addressing review questions	High
Weight of evidence D: Overall weight of evidence	Medium high

Wierstra R (1984) A study on classroom environment and on cognitive and affective outcomes of the PLON curriculum. *Studies in Educational Evaluation* 10: 272-282.	
Country	The Netherlands
Nature of interaction	Full context-based course
Name of course	Physics Curriculum Development Project (PLON)
Age range of learners	15 to 16
Science discipline	Physics
Type of study	Researcher-manipulated evaluation, non-randomised controlled trial
Aims of study	To investigate some aspects of inquiry-learning. Three domains of the effects of the MAVO (intermediate level) of the PLON curriculum were measured: cognitive achievement, affective and attitudinal outcomes, and the learning environment in the classroom. The purpose was to assess effects within each of these domains as well as the relationships between them.
Detail of sample	The sample was 297 boys and 101 girls in Dutch secondary schools. Nine classes (experimental) were following the PLON curriculum and six the non-PLON curriculum (control).
Methods to collect data	Individualized Environment Questionnaire, using ten items on a five-point scale to assess student perceptions of actual and preferred learning environment Attitude (enjoyment/relevance/affect) questionnaire of 12 items, using the five-point Likert scale Physics achievement tests from PLON and traditional physics exams
Methods used to analyse data	Correlations, t-tests and MANOVA
Summary of results	• In both the PLON group and the control group, it was found that enjoyment and relevance were correlated with the two achievement scores. • With regard to the cognitive outcomes, no differences were found in the total test. However, PLON pupils did just significantly better on test A (PLON based on PLON exams) and the control group did significantly better on test B (based on traditional physics exams. • In PLON classes, there is considerably more enquiry-learning than in control classes (as reported by pupils). • In both groups, it was found that a learning environment with more emphasis on inquiry-learning has a positive influence on the attitude to physics; moreover, in the PLON group, there were significant effects of inquiry learning (as expressed as class scores) on cognitive

	achievement. The greater amount of inquiry learning has a more favourable affective outcome for the PLON group. • Gender also had an impact. Boys showed higher scores on enjoyment and both achievement tests. Girls had a slightly more positive belief in learning from investigation. • Pupils' preference for inquiry learning is strongly influenced by the actual amount of inquiry learning he or she experiences.
Conclusions in relation to understanding	'In spite of the more positive attitudes of PLON pupils to physics and in spite of the positive correlations between attitude and achievement, there was no evidence for a better achievement by PLON pupils..' (p 280)
Conclusion in relations to attitude	'We found large and significant differences between PLON classes and control classes with regard to the amount of inquiry-learning as measured by the perceptions of pupils. However there were also large differences amongst PLON classes in the amount of inquiry-learning. This is because of the PLON philosophy of giving teachers a great deal of freedom to create their own class environment'. (p 280) 'Nevertheless, this study shows that the more inquiry based the classroom environment, the more favourable the pupils' attitude to physics (in both groups)'. (p 280)
Weight of evidence A: Trustworthiness of study findings in relation to answering study question	Medium
Weight of evidence B: Appropriateness of research design and analysis for addressing review question	Medium high
Weight of evidence C: Relevance of focus of study for addressing review question	High
Weight of evidence D: Overall weight of evidence	Medium high

Wierstra RF, Wubbels TA (1994) Student perception and appraisal of the learning environment: core concepts in the evaluation of the PLON physics curriculum. *Studies in Educational Evaluation* 20: 437-455.	
Country	The Netherlands
Nature of interaction	Full context-based course
Name of course	Physics Curriculum Development Project (PLON)
Age range of learners	15 to 16
Science discipline	Physics
Type of study	Researcher-manipulated evaluation, non-randomised controlled trial
Aims of study	To discover to what extent the PLON curriculum and in particular its reality- and activity-centeredness result in instruction that students experience as meaningful, and to find out how these curriculum characteristics affect affective and cognitive student outcomes. The study was based on a four-week teaching unit 'Traffic' which encompassed the teaching and learning of mechanics.
Detail of sample	Pupils in general, mixed sex, secondary education in schools in the Netherlands in 1988. The experimental group was made up of eight

	classes (totalling 209 pupils) taking the PLON physics curriculum and the control group made up of 16 classes (with 355 pupils) taking two traditional physics curricula (Middelink and Scoop).
Methods to collect data	• Students' cognitive outcomes data were collected by a 19-item conventional, multiple-choice standard physics tests covering the knowledge and performance common to the PLON and control curricula. • Data on affective outcomes (student appreciation of the unit, perceived instructiveness, general attitude to physics and student entering attitudes) were measured with Likert-type questionnaires. • Perceptions of the classroom environment were measured by a questionnaire on a five-point scale. • Details of pupil gender were also taken into account when scoring for cognitive outcomes.
Methods used to analyse data	Analyses of variance. Correlation analyses between various aspects measured were also carried out to check for validity and reliability.
Summary of results	• PLON students experienced the lessons as significantly more reality- and activity-centred than the control students. • Scores for PLON student appreciation and perceived instructiveness of mechanics lessons were significantly less than those for students on traditional courses. This was in contrast to expectations. • The influence of reality-centredness on the students' perceived instructiveness seems to be stronger than the influence of activity-centredness. • There was no significant difference between the two groups with respect to cognitive outcomes.
Conclusions in relation to understanding	'There appears to be no significant effect of the curriculum on the [cognitive] test scores. Unlike some critics' expectations, the PLON curriculum's emphasis on physics in everyday life was not paid for with lower-standard cognitive outcomes'. (p 451)
Conclusion in relations to attitude	'Our results in the evaluation of the unit Traffic indicate that reality-centeredness of the learning environment is a notable characteristic of the learning environment promoting perceived instructiveness. … Activity-centeredness also seems to promote affective outcomes, but more the appreciation of lessons than the perceived instructiveness. …In the study about Traffic we found that that the learning environment was perceived as more reality- and activity-centred in the PLON classes than in the control classes. Despite this difference and the correlations between reality- and activity-centredness and the affective outcomes, these outcomes tended to be lower in the PLON than the control classes'. (p 452)
Weight of evidence A: Trustworthiness of study findings in relation to answering study question	High
Weight of evidence B: Appropriateness of research design and analysis for addressing review question	Medium high
Weight of evidence C: Relevance of focus of study for addressing review question	High
Weight of evidence D: Overall weight of evidence	High

Yager RE, Weld JD (1999) Scope, sequence and co-ordination: The Iowa project, a national reform effort in the USA. ***International Journal of Science Education*** **21: 169-194.**	
Country	USA
Nature of interaction	Full context-based course
Name of course	Scope, Sequence and Co-ordination (SS&C): The Iowa Project
Age range of learners	12 to 14
Science discipline	Science
Type of study	Researcher-manipulated evaluation, non-randomised controlled trial.
Aims of study	To assess changes in teaching and student learning for those students following the Iowa-SS&C curriculum compared with students following a textbook course (Data was reported for teachers and students. This review has considered the student component only.) Of particular interest to the SS&C evaluation team were the six domains of concept, process, creativity, attitude, applications and connections, and world view. Career-orientations were also investigated.
Detail of sample	• The main sample was students from across five school districts; 5,270 students in the total experimental group (SS&C) and 1,320 in the total control (non-SS&C) group. Students were allocated to the two categories by class sections and care was taken that these were representative of the demographic composition of the Iowa districts areas involved (data given). • A subset of 364 experimental and 359 control students were involved in the career-orientation comparison.
Methods to collect data	• Test instruments for the curriculum-based evaluation were taken from the *Iowa Assessment Package*. The four domains of process, creativity, attitude and world view were pre- (September) and post- (April/May) tested. Concept and application were not pretested but an 'improvement rate' formula was used to provide scores for analysis. • Career-orientation was determined by a four-point scale questionnaire 'very highly agree' to 'do not agree'.
Methods used to analyse data	Analysis of covariance (ANCOVA)
Summary of results	• Students achieve better results in SS&C courses than they did in traditional text courses in all six achievement areas: concept, process, application, creativity, attitude and world view. • SS&C courses tended to be particularly useful to females and students of low ability. • Students following the SS&C course showed a significantly increased positive attitude to science as a career in four out of five questions whereas non-SS&C students only showed a significantly more positive attitude to science as a career in one of the five questions.
Conclusions in relation to understanding	• 'Students in IOWA-SS&C classrooms grow in concept and process skill mastery, applications of both in new situations, creativity skills (questioning, identifying causes, predicting consequences). Low ability and female students are especially well served in SS&C classrooms'. (pp 185-186)
Conclusion in relations to attitude	• 'Students in IOWA-SS&C classrooms grow in … improved attitudes to science study, classes, teachers and careers. Low ability and female students were especially well served in SS&C classrooms'. (pp 185-186) • Students of both sexes following the SS&C course showed a significantly increased positive attitude to science as a career compared to non-SS&C students.
Weight of evidence A:	High

Trustworthiness of study findings in relation to answering study question	
Weight of evidence B: Appropriateness of research design and analysis for addressing review question	Medium high
Weight of evidence C: Relevance of focus of study for addressing review question	High
Weight of evidence D: Overall weight of evidence	High