REVIEW

October 2005

A meta-analysis of the impact of the implementation of thinking skills approaches on pupils

Review conducted by the Thinking Skills Review Group

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

UNIVERSITY OF NEWCASTLE

AUTHORS

This report was written by a core review team from the Centre for Learning and Teaching based at the University of Newcastle upon Tyne, namely Steve Higgins, Elaine Hall, Viv Baumfield and David Moseley, with other members of the Thinking Skills Review Group and the Consultation and Critical Advisory Panel providing support, guidance and advice. Other members of Thinking Skills Review Group and Consultation and Critical Advisory Panel provided valuable support and critique influencing both the direction and scope of the review, and the reviewing of drafts of the report.

This document reports the details of a quantitative meta-analysis of the impact of thinking skills approaches on pupils' attainment and attitudes in schools. It sets out the background, processes and findings of the review. The Review Group is co-ordinated by the Centre for Learning and Teaching at Newcastle University and was undertaken by a core team, including teachers, local education authority (LEA) advisers, and academics. Practitioners played a key part in this review, particularly in defining the scope of the review, and clarifying the terms used in searching and keywording so as to identify studies with practical implications for teachers and schools.

REVIEW GROUP MEMBERSHIP

Dr Vivienne Baumfield	Senior Lecturer in Education, University of Newcastle
Marie Butterworth	Deputy Headteacher, Heaton Manor School
Graham Downey	Headteacher, Kramel First School, Northumberland
Maggie Gregson	Lecturer in Post-compulsory Education, University of Sunderland
Elaine Hall	Research Associate, University of Newcastle
Steve Higgins	Senior Lecturer in Education, University of Newcastle
Dr Mei Lin	Lecturer, University of Newcastle
David Moseley	Reader in Educational Psychology, University of Newcastle
Mel Rockett	LEA adviser, Northumberland

CONSULTATION AND CRITICAL ADVISORY PANEL

Sue Eagle	Primary School Headteacher, Norfolk
Jane Brooke	LEA Adviser, Cheshire
Dr Carol McGuinness	Professor of Educational Psychology, Queen's University, Belfast
David Thacker	Secondary Headteacher (retired)
Dr Iddo Oberski	Lecturer in Education, Stirling University
Dr Carolyn Tan	Senior Lecturer in Early Years Education, NIE, Nanyang Technological University (NTU), Singapore
Dr William Wu	Co-Director of the Thinking Qualities Initiative, Centre for Educational Development, Hong Kong Baptist University

CONFLICTS OF INTEREST

Throughout the review, we have tried to be consistent and transparent in the process of conducting the review, working within the EPPI-Centre guidelines, methodology and quality-assurance procedures for systematic reviewing, and involving members of the Review Group and Advisory Panel. We wanted to ensure that our own interest in the implementation and evaluation of thinking skills programmes and approaches did not influence our working processes or findings. We can, however, record our keen interest in the outcomes of the review and acknowledge that this may have influenced the review in ways which are not apparent to us.

ACKNOWLEDGEMENTS

This group is registered as an 'EPPI-Centre Review Group' and is part of the evidence-informed initiative in education funded by the Department for Education and Skills (DfES) at the EPPI-Centre (Evidence for Policy and Practice Information and Co-ordinating Centre), Social Science Research Unit, Institute of Education, University of London. The training and support from the EPPI-Centre team was invaluable in conducting the review, and the help and advice from Jo Garcia, Zoe Garrett, Angela Harden, Nicholas Houghton and James Thomas were especially helpful.

The group also acknowledges the feedback of referees in developing the thinking of the group. Comments on the initial proposal and on the protocol were useful in sharpening aspects of the review. Comments from reviewers on an earlier draft of this report are also gratefully acknowledged.

Funding

The funding provided by the EPPI-Centre was essential to establish the group and cover the costs of attending training. We estimate this funding covered a little less than one-third of the actual costs of the review. Additional costs have been met by the Centre for Learning and Teaching at Newcastle University and academic staff have benefited from funding for research from the Higher Education Funding Council for England (HEFCE). The support of the schools and local education authorities (LEAs) as well as the individual themselves in the Review Group and Consultation and Critical Advisory Panel was essential. Without this generous donation of time and effort, the review could not have been undertaken.

LIST OF ABBREVIATIONS

BDI	Biblioscape Database
BEI	British Education Index
CASE	Cognitive acceleration through science education
CERUK	Current Educational Research in the UK
DfES	Department for Education and Skills
ECO	Electronic Collections Online
EPPI-Centre	Evidence for Policy and Practice Information and Co-ordinating Centre
ERA	Education research abstracts
ERIC	Educational Resources Information Centre
FIE	Feuerstein's instrumental enrichment
HEI	Higher education institution
IBSS	International Bibliography of the Social Sciences
LEA	Local education authority
OCLC	Online Computer Library Centre
OFSTED	Office for Standards in Education
REEL	Research Evidence in Education Library
TTA	Teacher Training Agency

This report should be cited as: Higgins S, Hall E, Baumfield V, Moseley D (2005) A meta-analysis of the impact of the implementation of thinking skills approaches on pupils. In: *Research Evidence in Education Library*. London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.

TABLE OF CONTENTS

SUMMARY 1

Background..1
Aims and review questions 2
Methods 2
Results...3
Conclusions 3
Implications..4
Strengths and limitations 5

1. BACKGROUND 6

1.1 Aims and rationale for current review 6
1.2 Definitional and conceptual issues 6
1.3 Policy and practice background 8
1.4 Research background 8
1.5 Authors, funders and other users of the review 11
1.6 Review questions..11

2. METHODS USED IN THE REVIEW 12

2.1 User-involvement..12
2.2 Identifying and describing studies..12
2.3 In-depth review 13
2.4 Synthesis of evidence 14
2.5 In-depth review: quality-assurance process 16

3. IDENTIFYING AND DESCRIBING STUDIES: RESULTS 18

3.1 Origin of studies for the meta-analysis 18
3.2 Characteristics of the included studies 19
3.2 Characteristics of the included studies 20

4. Meta-analysis..23

4.1 Study quality and weight of evidence 23
4.2 Effect sizes and publication bias...24
4.3 Study heterogeneity...26
4.4 Quantitative synthesis of evidence 28
4.5 In-depth review: quality-assurance results 33
4.6 Nature of actual involvement of users in the review and its impact 33

5. FINDINGS AND IMPLICATIONS..34

5.1 Summary of principal findings...34
5.2 Strengths and limitations of this systematic review 35
5.3 Implications..37

6. REFERENCES 39

6.1 Studies included in map and synthesis...39

6.2 Other references used in the text of the report 41

Appendix 1.1: Consultation and Critical Advisory Panel 46

Appendix 2.1: Inclusion and exclusion criteria 47

Appendix 2.2: Search strategy for electronic databases 48

Appendix 2.3: EPPI-Centre keyword sheet, including review-specific keywords 50

Appendix 2.4: Calculating Hedges' *g* 52

Appendix 4.1: Summary of effects 54

Appendix 4.2: Details of studies included in the in-depth review 59

SUMMARY

Background

The teaching of thinking skills is an explicit part of the National Curriculum in England and Wales, and contributes directly to Department for Education and Skills (DfES) initiatives such as *Teaching and Learning in the Foundation Subjects* (DfES, 2004a) and *Leading in Learning* at Key Stage 3 (DfES, 2005) which emphasise the importance of thinking skills approaches for the promotion of effective questioning and extending pupils' oral responses in classrooms, as well as the potential contribution to assessment for learning. Thinking skills are also an important part of the developing Primary National Strategy aims (DfES, 2004b). However, thinking skills do not form a discrete and explicit programme of study and appear unevenly in the different subjects of the National Curriculum, which makes it challenging for schools to ensure progression in their teaching.

Our working definition for the purposes of this review is that thinking skills interventions are approaches or programmes which identify for learners translatable, mental processes and/or which require learners to plan, describe and evaluate their thinking and learning. These can therefore be characterised as approaches or programmes which:

- require learners to articulate and evaluate specific learning approaches; and/or
- identify specific cognitive, and related affective or conative processes that are amenable to instruction.

This definition is consistent with the definition used to identify and analyse thinking skills frameworks and taxonomies in other work undertaken by the Centre for Learning and Teaching (for example, Moseley *et al*., 2004, 2005a, 2005b).

A thinking skills approach therefore not only specifies *what* is to be taught but also *how* it is taught: the content of lessons and the teaching approach form an integral part of thinking skills approaches to teaching and learning. Examples of programmes and approaches commonly used in schools are instrumental enrichment (Feuerstein *et al.*, 1980), philosophy for children (Lipman *et al.*, 1980), cognitive acceleration through science education (CASE) (Adey *et al.*, 1995), and Somerset thinking skills (Blagg *et al.*, 1988). Considerable interest has also been shown by teachers and policymakers in how these more formal programmes can be integrated effectively or 'infused' into teaching approaches and adopted more widely by teachers (Leat and Higgins, 2002; McGuinness, 1999; McGuinness *et al.*,1995).

A meta-analysis was needed for the following reasons:

- to provide potential users with an estimate of the relative impact of thinking skills interventions, thus extending the scope of earlier reviews which have attempted to evaluate evidence from a range of thinking skills approaches (for example, Sternberg and Bhana, 1986); or which have focused on a particular programme (such as Romney and Samuels' (2001) meta-analysis of evidence of the impact on learners of Feuerstein's instrumental enrichment);

- to quantify the impact of thinking skills interventions in order to test the conclusions of the mainly positive but descriptive reviews in the UK (for example, Higgins *et al*., 2004; McGuinness, 1999; Wilson, 2000);

- to compare the impact of thinking skills interventions with other researched educational interventions (for example, Hattie *et al.*, 1992; Marzano, 1998; Sipe and Curlette, 1997).

In our first review (Higgins *et al.*, 2004), we identified and described 191 studies available up until 2002. We used narrative synthesis methods to address the question of what the impact of thinking skills interventions is on pupils. Twenty-three studies were included and reviewed in-depth following EPPI-Centre reviewing guidelines (EPPI-Centre, 2001, 2002).

This review concluded that the selection and implementation of thinking skills approaches needed to be based on more precise information on their effectiveness and efficiency. Meta-analysis is a method for pooling the quantitative estimates of effects of interventions from multiple studies to give a more reliable and precise estimate of their benefits (or potential harm). Comparing these estimates across different types of interventions can also pinpoint which aspects of interventions offer the most potential in the classroom. Meta-analysis is proving to be a useful approach to addressing the key question of practitioners interested in thinking skills in terms of ‘what works’ in education (for example, Hattie *et al*. 1996; Marzano *et al.* 2001; Sipe and Curlette, 1997).

Aims and review questions

The overall aim of the Thinking Skills Review Group is to investigate the impact of thinking skills interventions on teaching and learning in classrooms over a series of focused reviews.

Our main review question is as follows:

What is the impact of the implementation of thinking skills interventions on teaching and learning?

For this review, a narrower focus was identified for the central question about the quantitative impact of thinking skills interventions on pupils to provide a quantitative synthesis of evidence in this area:

What is the quantitative evidence for impact on pupils’ attainment and attitudes in schools?

Methods

Relevant studies in the area of thinking skills were obtained by systematically searching a number of online databases of educational research literature, by identifying references in reviews and other relevant books and reports, and from contacts with expertise in this area. Twenty-six of the studies identified for this review were obtained from the database which resulted from the first thinking skills review (Higgins *et al.*, 2004); a further three resulted from updating the original search and applying the more stringent criteria required for a quantitative synthesis.

Studies were selected for the meta-analysis if they had sufficient quantitative data to calculate an effect size (relative to a control or comparison group of pupils) and if the number of research subjects was greater than 10. Effect sizes were calculated from the reported data and combined statistically using quantitative synthesis.

Results

Twenty-nine studies were identified which contained quantitative data on pupils' attainment and attitudes suitable for meta-analysis. The studies come from a range of countries around the world with half set in the US and UK. The studies broadly cover the ages of compulsory schooling (5–16) and include studies set in both primary and secondary schools. A number of named thinking skills interventions are included, such as Feuerstein's instrumental enrichment (FIE) and cognitive acceleration through science education (CASE) as well as studies which report a more general thinking skills approach (such as the development of metacognitive strategies).

The quantitative synthesis indicates that thinking skills programmes and approaches are effective in improving the performance on tests of cognitive measures (such as Raven's progressive matrices) with an overall effect size of 0.62. (This effect would move a class ranked at 50th place in a league table of 100 similar classes to 26th or a percentile gain of 24 points.) However, these approaches also have a considerable impact on curricular outcomes with the same effect size of 0.62. The overall effect size (including cognitive, curricular and affective measures) was 0.74.

Conclusions

Overall, the quantitative synthesis indicates that, when thinking skills programmes and approaches are used in schools, they are effective in improving pupils' performance on a range of tested outcomes (relative to those who did not receive thinking skills interventions). The magnitude of the gains found appears to be important when compared with the reported effect sizes of other educational interventions.

This review found an overall mean effect of 0.62 for the main (cognitive) effect of each of the included the studies, larger than the mean of Hattie's vast database of meta-analyses at 0.4 (Hattie, 1999) but very similar to the overall figure reported by Marzano (1998, p 76) of 0.65 for interventions across the knowledge, cognitive, metacognitive and self-system domains. In particular, our study identified metacognitive interventions as having relatively greater impact, similar to Marzano's study.

Looking at a smaller part of our review, Feuerstein's instrumental enrichment is one of the most extensively researched thinking skills programme. Our results broadly concur with those of Romney and Samuels (2001), whose meta-analysis found moderate overall effects and an effect size of 0.43 on reasoning ability (p 28). Our findings were of the same order, with an overall effect size of 0.58 (one main effect from each of seven studies included) and an effect size of 0.52 on tests of reasoning (one main effect from four studies).

There is some indication that the impact of thinking skills programmes and approaches may vary according to subject. In our analysis there was relatively greater impact on tests of mathematics (0.89[1]) and science (0.78), compared with reading (0.4).

Implications

Thinking skills programmes have been extensively used around the world for a number of years. There, a growing body of accumulated evidence that they are effective at improving pupils' performance on cognitive and curriculum tests when they are researched in school settings. Their effect is relatively greater than most other researched educational interventions. This review strengthens this evidence base.

For practitioners, thinking skills programmes and approaches are likely to improve pupils' learning. Their use in schools should therefore be supported. Some caution is required as there is some variation in the impact of such approaches according to subject, age and gender. This suggests that their use needs to be matched to the particular teaching context and monitored critically to ensure potential benefits.

For policy-makers, thinking skills programmes and approaches are an effective way to improve teaching and learning, and their use in schools should be encouraged. However, as it is not clear to what extent the benefits are due to specific aspects of the content of the programmes and their implementation or the changes in teaching and learning which ensue, it is not possible to provide precise recommendations.

Further research is needed to clarify the particular causes of the benefits and where thinking skills programmes and approaches have most impact (such as on different age groups or in different areas of the curriculum). In particular, the impact of thinking skills programmes and approaches on teaching and learning processes needs to be related to improvements in outcomes to identify the means by which the impact occurs.

Researchers and journal editors should note that studies which looked like they were relevant to the review were often excluded because basic information, such as number of pupils involved, was not included in the published papers. Moreover, the details of sampling strategies and full sets of results were frequently omitted. Abstracts sometimes referred to data which was not then reported in detail. Journals which used structured abstracts were more likely to contain more accurate and more complete information to support the systematic reviewing process.

[1] A percentile gain of 31 points

Strengths and limitations

Strengths

- The extent of the literature included in the review process, building on the mapping of this literature and narrative synthesis completed for the first review

- The use of meta-analysis to provide a quantitative synthesis of the research literature and an overall estimate of the impact of such approaches and the interpretation of this quantitative synthesis which contextualises thinking skills research within the broader field of educational research

- The close involvement of practitioner users groups in setting and refining the questions, and interpreting and disseminating the findings

Limitations

- Studies often reported little about the programmes themselves or aspects of their implementation and use in classrooms (such as changes in teaching and learning processes). It is therefore difficult to draw conclusions about any common features of programmes and approaches which may account for the positive impact reported.

- The review used a broad definition of thinking skills for its focus. As a result, there was considerable statistical heterogeneity in the results of the studies which indicates that caution is required in combining the effects and interpreting the findings.

- We were only able to identify and synthesise 29 studies within the timescale and resources for the review. A larger number of studies would enable further analysis (such as by age or subject) to make more specific recommendations for practitioners and policy-makers.

- Meta-analysis, or quantitative synthesis, is subject to a number of limitations and criticisms; this review is therefore open to such critique (see, for example, Chambers, 2004; Kulik and Kulik, 1989a).

1. BACKGROUND

1.1 Aims and rationale for current review

The overall aim of the Thinking Skills Review Group is to investigate the impact of thinking skills interventions on teaching and learning in classrooms over a series of focused reviews.

Our main review question is as follows:

What is the impact of the implementation of thinking skills interventions on teaching and learning?

For this review, a narrower focus was identified for the central question about the quantitative impact of thinking skills interventions on pupils:

What is the quantitative evidence for impact in schools on pupils' attainment and attitudes?

In our first review (Higgins *et al.*, 2004), we identified and described 191 studies available up until 2002. We used narrative synthesis methods to address the question of what the impact of thinking skills interventions is on pupils. Twenty-three studies were included and reviewed in-depth following EPPI-Centre reviewing guidelines (EPPI-Centre, 2001, 2002).

This first review concluded that the selection and implementation of thinking skills approaches needed to be based on more precise information on their effectiveness and efficiency. Meta-analysis is a method for pooling the quantitative estimates of effects of interventions from multiple studies to give a more reliable and precise estimate of their benefits (or any detrimental effects). Comparing these estimates across different types of interventions can also pinpoint which aspects of interventions offer the most potential in the classroom. Meta-analysis is proving to be a useful approach to addressing the key question of practitioners interested in thinking skills in terms of 'what works' in education (for example, Hattie *et al*., 1996; Marzano *et al.*, 2001; Sipe and Curlette, 1997).

A meta-analysis of the quantitative impact of the implementation of thinking skills approaches on pupils' attainment and attitude was therefore identified as a way to provide potential users with an estimate of the relative impact of thinking skills interventions by offering data to compare the impact of thinking skills interventions with other researched educational interventions (for example, Hattie *et al.*1992; Marzano, 1998). This quantitative synthesis would test the conclusions of the mainly positive but descriptive reviews in the UK (e.g. McGuinness, 1999; Wilson, 2000), Europe (e.g. Hamers and van Luit, 1999) and the US (e.g. Cotton, 1991).

1.2 Definitional and conceptual issues

The teaching of thinking skills is an explicit part of the National Curriculum in England and Wales, and contributes directly to Department for Education and Skills (DfES) initiatives, such as *Teaching and Learning in the Foundation*

Subjects (DfES, 2004a) and *Leading in Learning* at Key Stage 3 (DfES, 2005), which emphasise the importance of thinking skills approaches for the promotion of effective questioning and extending pupils' oral responses in classrooms as well as the potential contribution to assessment for learning. Thinking skills are also an important part of the developing Primary National Strategy aims (DfES, 2004b). However, thinking skills do not form a discrete and explicit programme of study and appear unevenly in the different subjects of the National Curriculum which makes it challenging for schools to ensure progression in their teaching.

The descriptive review by Carol McGuinness (1999) for the DfES provides an overview of current research into the teaching of thinking skills and builds on the work of earlier reviews in this area. Nisbet and Davies (1990) list 30 specific programmes and indicated that there were then over 100 on the market in America. Hamers and Van Luit (1999) show that this is not an English-speaking phenomenon and that interest in teaching thinking is evident among practitioners and educational researchers in many other European countries.

Thinking skills initiatives have been used in schools in the UK since the early 1980s and they have been in existence for somewhat longer, but the term itself is ambiguous and there is disagreement about how it relates more broadly to aspects of pedagogy. Our working definition for the purposes of this review is that thinking skills interventions are approaches or programmes which identify for learners translatable mental processes and/or which require learners to plan, describe and evaluate their thinking and learning. These can therefore be characterised as approaches or programmes which:

- require learners to articulate and evaluate specific learning approaches
- identify specific cognitive (and related affective or conative processes) that are amenable to instruction

A thinking skills approach therefore not only specifies the content of what is to be taught (often framed in terms of thinking processes, such as understanding, analysing or evaluating) but also the pedagogy of how it is taught (usually with an explicit role for discussion and articulation of both the content as well as the process of learning or metacognition). Implicit in the use of the term is an emphasis on so-called 'higher-order' thinking, drawing on Bloom and colleagues' taxonomy (Bloom *et al.,* 1956). This consists of six major categories arranged in the following order: *knowledge, comprehension*, *application*, *analysis*, *synthesis* and *evaluation*. The relationship among the categories along the continuum was presumed to constitute a cumulative hierarchy. (For further details about frameworks and taxonomies of thinking skills, see Moseley *et al*., 2004, 2005a, 2005b).

Examples of programmes and approaches commonly used in schools are instrumental enrichment (Feuerstein *et al*., 1980), philosophy for children (Lipman *et al.*, 1980), cognitive acceleration through science education (Adey *et al.*, 1995; Shayer and Adey, 2002), and Somerset thinking skills (Blagg *et al.*, 1988). Nickerson *et al.* (1985) attempted to impose a structure on these programmes by classifying them into five categories, a classification accepted by Garnham and Oakhill (1994), and Hamers and Van Luit (1999), although the former authors accepted that these were only broad categories.

1.3 Policy and practice background

Thinking skills approaches are generally popular with teachers (Baumfield and Oberski, 1998) and there is evidence that they seem to support changing patterns of interaction in classrooms (Higgins and Leat, 1997; Leat and Higgins, 2002). This understanding is influenced by concepts and ideas derived from cognitive acceleration (Adey and Shayer, 1994), Instrumental Enrichment (Feuerstein *et al*., 1980), Philosophy for Children (Lipman, 1991, 2003), 'probes' for understanding (White and Gunstone, 1992), reciprocal teaching (Palincsar and Brown, 1984), scaffolding and social constructivism (Wood and Wood, 1996), research on classroom talk (Edwards and Westgate, 1987; Mercer 1995), self-theories (Dweck, 1999), formative assessment (Black and Wiliam, 1998; Torrance and Pryor, 1998) and collaborative group work (Galton *et al*., 1999; Webb and Farrivar, 1994). This work has been used in research and development work with trainee and practising teachers, as a means by which teachers could put into practice or 'enact' findings from educational research (Higgins, 2001; Higgins and Moseley, 2002; Leat and Higgins, 2002). In England and Wales, thinking skills approaches have been influential in the development of the National Curriculum (McGuinness, 1999), and the development of the Key Stage 3 and Primary National Strategies.

1.4 Research background

There is a range of research evidence about whether thinking skills approaches and the underpinning metacognitive techniques are effective in raising pupils' attainment (e.g. Adey and Shayer, 1994; Romney and Samuels, 2001; Rosenshine and Meister, 1994; Trickey and Topping, 2004; Wong *et al*., 1985). The descriptive reviews by Carol McGuinness (1999) for the Department for Education and Employment and by Wilson (2000) for the Scottish Executive provide an overview of recent research into the teaching of thinking skills relevant to the UK. Hamers and Overtoom (1997), and Hamers and Van Luit (1999) summarise recent research in Europe. Most of these reviews are not explicit about their scope and scale.

Several meta-analyses have been conducted in the broader area of thinking and learning. Such techniques make it possible to compare the impact of different types of educational interventions using statistical techniques involving effect sizes. A few of these have focused on specific approaches rather than thinking skills more generally, and this review hopes to provide some answers about thinking skills more broadly conceptualised. This will also help to place thinking skills approaches in the wider context of educational research. Two notable studies have looked at a range of interventions and included aspects of thinking skills as part of their reviews. Hattie *et al.* (1996) evaluated the effects of a wide range of learning skills interventions on students' learning. A meta-analysis by Marzano (1998) was even broader in scope and larger in scale; it categorised all studies in terms of type of intervention and area of thinking affected, although it should be noted that it has not been subject to peer review but is available as a report posted on the internet.

These studies are consistent in finding that thinking techniques designed to be used by students and develop their self-regulation led to significantly better results than those designed to be presented by teachers. Although there was enormous diversity in the intervention studies selected by Marzano, ranging from a focus on specific skills (such as memorisation) to the use of disposition-monitoring strategies, he made the following claim about the importance of making aspects of thinking explicit or metacognition:

> Instructional techniques that employed the metacognitive system had strong effects whether they were intended to enhance the knowledge domains, the mental process within the cognitive system, the beliefs and processes within the self-system, or the processes within the metacognitive system itself. (p 127)

Overall, Marzano found that interventions which engage either the ‘self-system’ or the ‘metacognitive system’ lead to better knowledge outcomes (by six and five percentile points respectively) compared with those which are directed only at the use of cognitive skills. Nevertheless, there are some types of very effective intervention at the cognitive skill level. These are interventions which address experimental enquiry, using analogies, comparing and contrasting, idea representation, and the storage and retrieval of knowledge.

Sipe and Curlette (1997) undertook a ‘meta-synthesis’ or an analysis of 103 meta-analyses of student achievement using Walberg’s (1984) educational productivity model of student aptitude, instruction and environment as a framework. They found a mean effect of 0.375 for curriculum interventions. The interventions with the greatest effects were vocabulary instruction (1.15), accelerative instruction (0.88), mastery learning (0.82), direct instruction (0.82) and note-taking (0.71). Two of the interventions with the lowest impact were ability grouping (–0.038) and matched teacher/student cognitive style (0.03). They concluded that ‘the most effective curriculum interventions had an effect size close to or greater than .8’ and ‘the least effective curriculum interventions had effect sizes near zero and less than .2’ (p 653).

An example of a meta-analysis of an approach which fits our broader definition of thinking skills is Rosenshine and Meister’s (1994) review of reciprocal teaching. This is a teaching approach which features ‘guided practice in applying simple concrete strategies to the task of text comprehension’ (Brown and Palincsar, 1989). It includes cognitive techniques, such as summarisation, question generation, clarification and prediction supported through dialogue between teacher and students (or students and students) as they attempt to gain meaning from a text. Rosenshine and Meister’s review includes 16 quantitative studies of reciprocal teaching and found an average effect size of 0.32 when the impact of the intervention was measured using standardised tests, and an average effect size of 0.88 when more specific tests developed by the researcher were used.

An example of a meta-analysis of a specific thinking skills programme is Romney and Samuels (2001) review of Feuerstein’s instrumental enrichment (FIE). Proponents of FIE claim that the programme results in an improvement in school achievement, cognitive ability and classroom behaviour. However, because some outcome studies have produced negative results, Romney and Samuels undertook a meta-analysis in order to provide a more reliable and comprehensive assessment of the efficacy of FIE. A total of 40 controlled studies, comprising 47 different samples, were examined. Significant, although modest, average effect sizes were found in all three areas – achievement, ability, and behaviour – with the most extensive improvement being made in ability. Gains in spatial/perceptual ability were related to the length of the intervention (number of hours), and self-esteem was related to age, with older children showing increases and young children showing decreases. This provides powerful evidence for those considering using such a programme in schools. Not only does it suggest that gains, although modest, are likely to be achieved when using FIE, it also provides some pointers about how to implement the programme and some implications about attending to the impact on self-esteem of younger learners in particular. The average effect sizes are lower that those found in other programmes, but that might also be explained by the predominant use of FIE with pupils with special educational needs.

Another example of a meta-analysis of a particular thinking skills approach is Trickey and Topping's (2004) review of 'philosophy for children'. It provides a systematic and critical review of controlled outcome studies where the *Philosophy for Children* approach has been used in primary and secondary schools. Ten studies met their inclusion criteria as follows: measuring outcomes with norm-referenced tests of reading, reasoning, cognitive ability, curriculum measures or measures of self-esteem, child behaviour or pupil and teacher questionnaires. All the studies showed some positive outcomes with a mean effect size of 0.43. They concluded that *Philosophy for Children* has 'a consistent moderate positive effect… on a wide range of outcome measures'. This review further points to an additional benefit of meta-analysis which is the possible calculation of cost-benefit, arguing that, approaches such as *Philosophy for Children* are good value for money with a reliable (if moderate) effect size for relatively little investment. Although beyond the scope of this review, such analysis is clearly valuable for both policy-makers and practitioners in relating the potential benefits to the economic (or other) costs associated with their implementation.

In this review, we add to the contributions of previous systematic reviews by including a wider field of studies (i.e. more than a review of a single programme) and by identifying empirical classroom investigations with ecological validity (Gall *et al.*, 1996) for closer analysis. The studies have then been subjected to a systematic and rigorous review procedure that enables users to form a judgement on the relevance of approaches reported in terms of the review question.

A further quantitative synthesis was therefore needed for the following reasons:

- to quantify the impact of thinking skills interventions in order to test the conclusions of the mainly positive but descriptive or narrative reviews in the UK (for example, Higgins et al., 2004; McGuinness, 1999; Wilson, 2000)

- to provide potential users with an estimate of the relative impact of thinking skills interventions, thus extending the scope of earlier reviews which have attempted to evaluate evidence from a range of thinking skills approaches (e.g. Sternberg and Bhana, 1986); or which have focused on a particular programme (such as Romney and Samuels' (2001) meta-analysis of evidence of the impact on learners of Feuerstein's instrumental enrichment or Trickey and Topping's (2004) review of *Philosophy for Children*)

- to compare the impact of thinking skills interventions with other researched educational interventions (e.g. Hattie *et al.*1992; Marzano, 1998; Sipe and Curlette, 1997)

1.5 Authors, funders and other users of the review

The review set out to provide information for a range of audiences. The initial descriptive map and keywording of studies provide a valuable overview of the field for practitioners (particularly those interested in, or undertaking, classroom-based research), educational researchers themselves and those who fund research. The findings from this third phase of the review should enable these groups to find relevant studies about the implementation of thinking skills programmes and approaches which contain evidence of impact. The in-depth review and synthesis are intended to inform these audiences and summarise findings for policy-makers and practitioners more widely, with implications drawn out for policy and practice.

1.6 Review questions

The aim of this review is to investigate the impact of thinking skills interventions on pupils' attainment and attitude in school.

The main question for the review is as follows:

What is the quantitative impact of the use of thinking skills interventions on pupils' attainment and attitudes?

In order to address this question, attainment is narrowly conceptualised as performance on tests of:

- general thinking or cognitive functioning (including standardised tests, such as Raven's Progressive Matrices or tests devised by the researchers involved)

- curricular performance (standardised tests or school tests and examinations in specific subjects, such as mathematics, science, etc.)

Attitudes are perhaps less easily and certainly less frequently captured using quantitative measures; however, where reliable and valid measures were used, and where effect sizes were reported or could be calculated, these were included in the meta-analysis.

2. METHODS USED IN THE REVIEW

2.1 User-involvement

The wider Review Group includes representatives from key constituencies of users, such as practitioners from primary and secondary schools, LEA advisers and the research community. Two members of the core team were teacher researchers who had considerable experience of implementing and evaluating thinking skills approaches and were also familiar with research methods. We also used existing links with research centres in Singapore and Hong Kong to access an international perspective. The focus of the review was identified by members of the Review Group and Consultation and Critical Advisory Panel and refined in the light of comments from referees and EPPI-Centre staff. The outcomes of the first review were used to refine the review question.

2.2 Identifying and describing studies

2.2.1 Defining relevant studies: inclusion and exclusion criteria

The Review Group used the inclusion criteria from the first review and added further criteria for the meta-analysis (*in italics*). Studies were included which:

- were set in a school or schools and concerned with any section of the school population (including pupils with special needs)
- evaluated the impact of the implementation of thinking skills interventions on teaching and/or learning
- were concerned with the phases of compulsory schooling (5–16)
- contained empirical classroom research with data or evidence (pupil outcomes, classroom processes, teacher role)
- were written in English[1]
- reported quantitative data on pupils' achievement or attitudes
- used a control or comparison group
- reported pre- and immediate post-intervention measures[2]
- contained data on at least 10 pupils
- reported an effect size for the intervention OR

[1] It was beyond the funding of the review to translate papers from other languages. Although this may have restricted literature identified, every effort was made to identify studies from non-English-speaking countries, but published in English.
[2] There were too few studies which investigated retention or transfer effects with suitable data, so for comparability we only included studies with post-test data collected immediately after the intervention.

- contained sufficient data to calculate an effect size of the impact of the thinking skills intervention or approach – for example,

  means and standard deviations for pre- and post test scores for intervention and comparison group
  means, t-test values and group sizes
  one-way ANOVA for the groups (with F-values and sample sizes)
  one-way ANCOVA with error values and correlation (*r*) between the covariate and dependent measure.

2.2.2 Identification of potential studies: search strategy

The potential studies for this review were identified in two stages. First, the additional inclusion criteria were applied to the database of 191 reports from the first review (Higgins *et al.*, 2004). A further search was then undertaken by updating the original electronic searches and handsearches. The same search terms were applied to the same search gateways and databases for the years 2002 and 2003 (see Appendix 2.2). The combined inclusion criteria were then applied to these reports to identify potential studies for this reviews. In addition, any studies which had been obtained after the cut-off date for inclusion in the first review were considered.

2.2.3 Identifying and describing studies: quality-assurance process

The core review team moderated the use of the inclusion and exclusion criteria through meetings where members worked in pairs to apply the criteria on a sample of abstracts and full studies. A high level of consensus was reached in these meetings: inter-rater agreement was reached in 97% of cases on a sample of 60 abstracts undertaken at the end of this process.

2.3 In-depth review

2.3.1 Detailed description of studies in the meta-analysis

Detailed description was completed according to the EPPI-Centre guidelines for extracting data, and the software supported the reviewers in making a careful analysis of the content of the studies. We were particularly concerned with the ecological validity of the studies and so questions regarding how the intervention was implemented were important and cross-referenced with our review-specific keywords. Methodologically, studies were included if they reported a research design that involved the researcher planning or controlling the investigation to provide opportunities for before and after evaluation with comparison or control groups.

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

The EPPI-Centre guidelines for assessing the quality of studies in EPPI-Reviewer require the weight of evidence to be judged both according to the internal validity and reliability of each study, and the external or ecological validity in terms of the value for our particular review.

- Weight of evidence A refers to the internal consistency of the study in the sense of ‘Can the reported findings be trusted in answering the the researchers’ own study question?’ or the extent to which a study is carried out according to accepted practice for the methodology adopted.

- Weight of evidence B is concerned with the appropriateness or applicability of the research design for our review question.

- Weight of evidence C is concerned with the focus of the study for our review question.

- Weight of evidence D is concerned with the overall weight of evidence when A, B and C are combined.

- A, B, C and D are all classified as high medium or low. The classification of weight of evidence D is determined by the average grade given for A, B and C.

Issues in establishing the weight of evidence often revolved around the transparency of reporting and whether sufficient information was provided in the study to make judgements about aspects of the research, such as fidelity of implementation of the thinking skills programme or approach.

2.4 Synthesis of evidence

Information from those studies which addressed the questions was brought together within the conceptual and contextual framework introduced in Chapter 1. For this evaluation, the Review Group wanted to identify studies which reported quantitative data on the impact on pupils of the use of thinking skills programmes and approaches in schools, particularly where the usual teacher undertook the intervention with a normal class grouping (rather than a researcher with a particular group of pupils, for example) to ensure the ‘ecological validity’ of the study.

2.4.1 Statistical synthesis of outcomes (meta-analysis)

In the synthesis presented below (Chapter 4), the effectiveness of different thinking skills interventions was established by the effect size – a standard measure of impact for the range of outcomes in the included studies. This is a standardised measure which allows studies using different outcome measures to be compared using the same metric (i.e. the mean difference between groups divided by the standard deviation).

The advantage of translating the difference between intervention and control groups into this measure is that effect sizes can then be compared across studies that use very different dependent (test) measures. For example, in an analysis of 13 studies that produced 22 different effect sizes for the relationship between psychotherapy and asthma, Glass *et al.* (1981) report an average effect size of 0.85. One of the valuable aspects of effect size is that it is a measure of standard deviation units and can therefore be interpreted as a change in the percentile ranking of the 'average' subject in the intervention group. For example, the reported effect size of 0.85 is the same as saying that the average score of subjects in the intervention groups is 0.85 standard deviations above the mean of subjects in the control group. Therefore, the average pupil in the intervention group is at the 80th percentile of the control group or an increase of 30 percentile points. As a further indicator, Glass *et al.* (1981, p 102) suggest the example of an effect size of 1.0 corresponding to the difference of about a year of schooling as measured by achievement tests used in elementary or primary schools.

The effect sizes in this review (Hedges' *g*[1]) were calculated using the tools within EPPI-Reviewer which allows for the calculation of an effect size from a range of data reported such as t-test values or ANOVAs. The formulae for the equivalence of different effect sizes can be found in various meta-analysis texts, such as Lipsey and Wilson (2001) or Deeks *et al.* (2001).

An area of concern in this review (and for other similar reviews) is how to interpret the meaning of results presented in effect sizes. Cohen (1988) pointed out how his initial labeling of low (0.2), medium (0.5), and high (0.8) effects was somewhat arbitrary, and depended on what was being measured and how it was measured. Our approach was to be broadly comparative and to interpret effect sizes in terms of what is reported in the wider educational literature as well as to look at the relative impact of the studies within the review.

A meta-analysis essentially adds a number of studies together using a statistical method that gives the greatest weight to the studies with the smallest standard errors, which usually means the largest studies. It offered the best way to pool the results of a range of studies quantitatively so as to compare the impact of different thinking skills interventions with each other and to compare the impact of thinking skills approaches with other researched educational interventions (for example, Hattie *et al*., 1996; Marzano, 1998).

2.4.2 Publication bias

One potential weakness of systematic reviews is through publication bias. If studies showing a positive (beneficial) effect are more likely to be published than negative or inconclusive studies, this will give an inflated estimate of effect. One method of determining the existence of publication bias is to draw a funnel plot. This plots the effect size of a study (on the *x*-axis) against its sample size (on the *y*-axis). Very small studies will have a high probability of showing an inconclusive (not statistically significant) result, even if the intervention is effective, just as they will have a raised probability of showing a positive effect if the intervention is ineffective. If there is no publication bias, small studies should be scattered along the *x*-axis, with the larger studies being situated closer to the true estimate of effect (as it is assumed that they are less subject to variability). The result should

[1] Hedges' *g* is an inferential measure. It is normally computed by using the square root of the Mean Square Error from the analysis of variance testing for differences between the two groups. It can also be calculated from other measures of effect size.

therefore resemble an inverted funnel. We used a funnel plot to investigate whether or not there was any publication bias in the studies in the meta-analysis.

2.4.3 Fixed and random effects models of meta-analysis

Fixed-effect models of meta-analysis make a mathematical assumption that every study is estimating the same treatment effect or that the underlying treatment effects are identical. Random effects models of meta-analysis are an alternative approach to meta-analysis which are not based on this assumption that a common ('fixed') treatment effect exists. Instead, they assume that studies are estimating different, but related, treatment effects, with the differences between these represented by random variation. This assumption that the true treatment effects in the individual studies may be different from each other means there is no single number to estimate in the meta-analysis, but a distribution of numbers. The most common random effects model also assumes that these different 'true' effects are normally distributed. The meta-analysis therefore estimates the mean and standard deviation of the different effects. The meta-analysis software within EPPI-Reviewer uses DerSimonian and Laird's (1986) random effects method.

One indicator of the appropriateness of the different models is a test of statistical heterogeneity. If statistical heterogeneity is identified, it is normally recommended that a fixed effects model should not be used, or, at least, the results should be compared with those obtained with random-effects models.

The impact of interventions is always likely to vary by chance, but testing for heterogeneity investigates whether there is more variation than would be expected by chance alone.

2.4.4 Weight of evidence

The EPPI-Centre guidelines were used to establish the relative 'weight of evidence' for each of the included studies. This enabled us to investigate whether there was any link both between the quality of the study (in particular weight of evidence A: the study's quality in terms of its appropriateness to answer its own research questions) and between the overall weight of the study in terms of its appropriateness in answering the review question.

2.5 In-depth review: quality-assurance process

Data were double-entered onto EPPI-Reviewer (a web-based database) by at least two reviewers working independently. In cases where there was initial disagreement about data extraction or quality appraisal, this was discussed and resolved. Three members of the EPPI-Centre were involved in this process so as to ensure consistency across systematic reviews. Seven of the 30 studies (23%) were evaluated independently by EPPI-Centre staff, then the results were compared and any differences discussed and resolved. Inter-rater reliability was achieved in over 95% of coding, with most differences relating to the extra detail of what was recorded (rather than the coding itself) or inference from the reporting (e.g. country of study might be inferred from the authorship of the report).

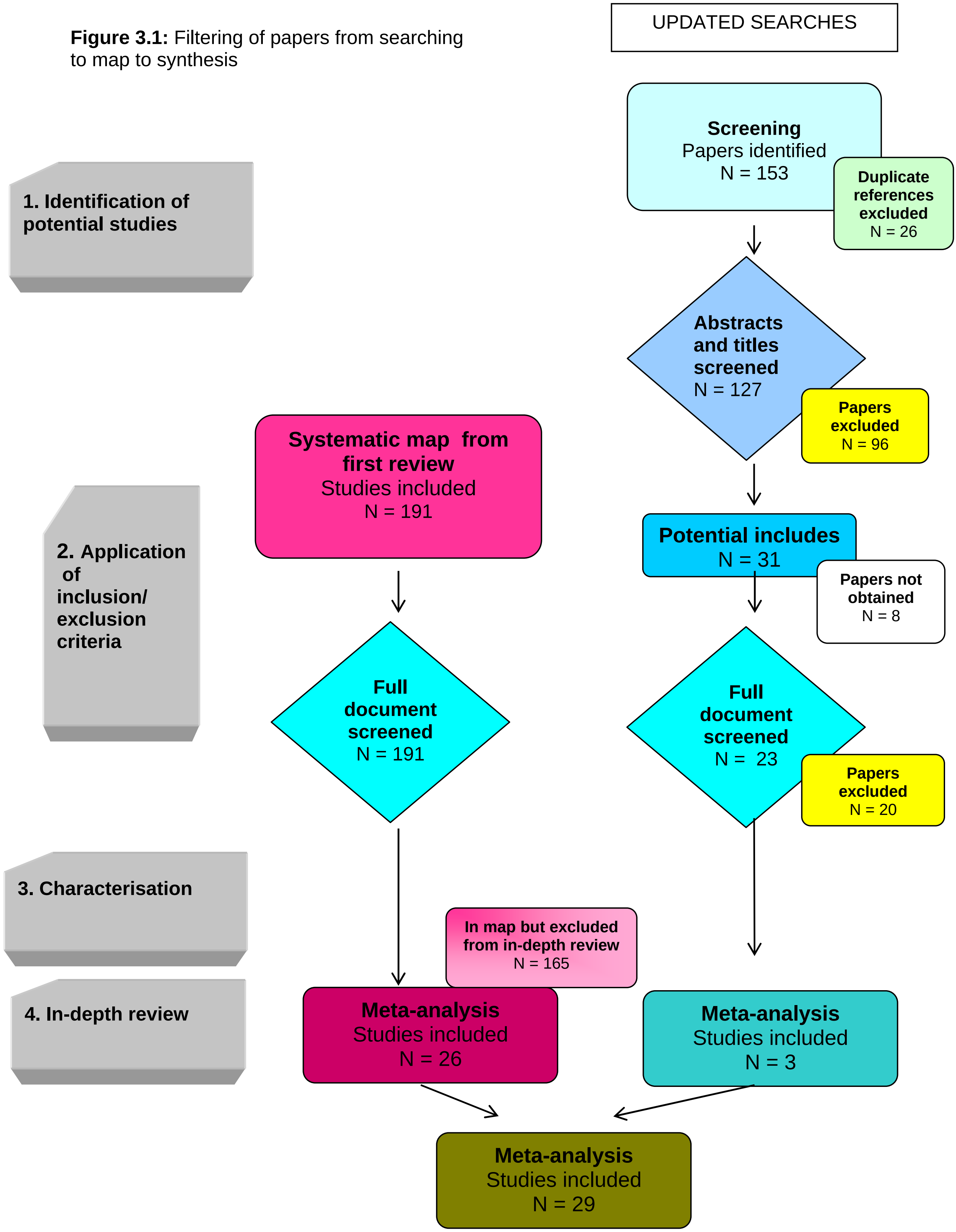
The synthesis of evidence was reviewed at a core review team meeting where the weighting of evidence was discussed and the relevance of studies for our key question debated and agreed. In particular, the issue of the applicability of the research to practice in schools was identified as a key issue as well as the challenge of identifying specific implications for teaching in schools.

3. IDENTIFYING AND DESCRIBING STUDIES: RESULTS

3.1 Origin of studies for the meta-analysis

Figure 3.1 shows how the studies for the meta-analysis were identified. From the studies in the Review Group's original map (Higgins *et al.*, 2004) 26 were identified as meeting the inclusion criteria for this meta-analysis. An additional search identified a relatively small number of further reports (127) which had been published or indexed since the initial search was undertaken. The inclusion criteria for this review were then applied to these reports to identify studies with quantitative data suitable for meta-analysis. Only three further studies met these more stringent criteria. In total, this gave 29 studies for the meta-analysis. Of the studies not obtained (either not available within the timescale for the review, or simply not available at all), the majority were older studies (pre-1980) or were reports and theses (particularly Masters' thesis). It is difficult to assess the impact of this on the review. Evidence from other reviews suggests that unpublished studies are likely to have a lower overall effect size (Lipsey and Wilson, 1993; Sipe and Curlette, 1997) and meta-analyses with fewer that 50 studies tend to report higher effect sizes (Cooper *et al*., 1995).

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



3.2 Characteristics of the included studies

3.2.1 Country of included studies

The studies in the meta-analysis come from a range of countries, with half from the US and UK and a range of other countries represented.

Table 3.1: Countries of studies in the in-depth review (N = 29)

Country	Number of studies
Australia	3[5]
Canada	2
Hungary	1
Israel	2
The Netherlands	1
Nigeria	1
Pakistan	1
The Philippines	1
Taiwan	1
South Africa	1
UK	7
USA	8[5]
Total (mutually exclusive)	***29***

3.2.2 Educational setting

The studies in the meta-analysis represent a range of educational settings.

Table 3.2: Educational setting of studies in the in-depth review (N = 29)

Educational setting	Number of studies
Primary school	9
Secondary school	20
Special needs school[6]	2
Total (not mutually exclusive)	***31***

3.2.3 Curriculum focus of the studies in the in-depth review

The majority of the studies in the meta-analysis focus on science, mathematics and aspects of literacy (mainly reading comprehension).

[5] Inference for one paper from affiliation of the author of the study.
[6] Both Special Needs schools were also secondary schools

Table 3.3: Curriculum focus of the in-depth studies (N = 29)

Curriculum focus	Number of studies
Citizenship	1
General/Cross-curricular/Other	7
ICT	1
Literacy – first languages	7
Mathematics	9
Science	9
Total (not mutually exclusive)	***34***[7]

3.2.4 Sample size

The distribution of sample size is shown in Table 3.5. The majority of studies analysed for the in-depth review have samples between 11 and 100 pupils.

Table 3.5: Distribution of sample size in the in-depth studies (N = 29)

Sample size (experimental and controls)	Studies
11–50	11
51–100	4
101–500	13
500+	1
Total (mutually exclusive)	***29***

Study type

Due to the nature of the review, all the studies contained comparison or control groups. Of these, the 13 allocated subjects to intervention and control groups randomly[8] or used existing groupings (e.g. classes of pupils) to allocate subjects (9). For seven of the studies, matched groups of pupils were identified in existing groupings in schools on the basis of pre-test scores (i.e. there was no prospective allocation into intervention or control groups).

Table 3.6: Allocation of research groupings (N = 29)

Study allocation	Studies
No prospective allocation	7
Random allocation	13
Non-random allocation	9
Total (mutually exclusive)	***29***

[7] Some studies reported on impact on more than one curriculum subject.
[8] Or pseudo-randomly using a computer generated sequence

Thinking skills programmes

The majority of studies included in the in-depth review are evaluations of specific named thinking skills programmes (22/29). Seven of the studies are evaluations of FIE and four are of evaluation of CASE or related approach. Five of the studies are explicit about targeting metacognitive strategies, although the majority involved teaching of metacognitive skills, at least implicitly, as part of the approach.

4. META-ANALYSIS

4.1 Study quality and weight of evidence

4.1.1 Quality of reporting

Most of the studies report clear aims, underpinning rationale and context. However, there are rarely details about the overall sampling strategy (particularly the sampling frame) by which schools were involved in the research. The studies also vary in terms of the extent to which they report on other aspects of the research, such as the classroom processes involved – for example, in ensuring that the thinking skills intervention enabled pupils to talk about their thinking – often referred to as 'fidelity of implementation' (Fullan and Pomfret, 1977). The majority of studies are clear about the quantitative analysis and methods used to ensure the transparency of reporting (such as detailing reliability and validity of measures used or the method of analysis chosen).

4.1.2 Weight of evidence

Information from those studies which address the question was brought together and the weight of evidence judgements for each of the studies in the in-depth review are shown in Table 4.1. (See section 2.3.2 for detail on how the weight of evidence judgements were made.)

Table 4.1: Weight of evidence table

Item ID	Author/date	A	B	C	D
IT14198	Adey and Shayer (1990)	H	H	H	H
IT16621	Adey *et al.* (2002)	H	H	H	H
IT13471	Cardelle-Elawar (1992)	H	H	M	H
IT11895	Chang and Barufaldi (1999)	H	H	M	H
IT16404	Collings (1994)	M	H	M	M
IT13508	Csapó (1992)	H	H	H	H
IT18183	Cunningham *et al.* (2002)	H	H	H	H
IT16665	De Koning and Hamers (1999)	M	M	M	M
IT16405	Donegan and Rust (1998)	H	M	M	M
IT15899	Greenberg (2000)	M	M	M	M
IT13879	Haywood *et al.* (1988)	M	M	L	M
IT16406	Hoek *et al.* (1999)	H	H	H	H
IT16647	Iqbal and Shayer (2000)	M	M	M	M
IT16476	Kaniel and Reichenberg (1992)	H	H	H	H
IT13839	Kramarski and Mevarech (1997)	H	M	M	M
IT16087	Maqsud (1998)	H	H	M	H
IT16229	Martin (1984)	L	L	L	L
IT13613	Mercer *et al.* (1999)	L	L	M	L

Item ID	Author/date	A	B	C	D
IT16478	Muttart (1984)	L	L	L	L
IT16407	Naval-Severino (1993)	M	M	M	M
IT16054	Oladunni (1998)	H	H	H	H
IT16408	Riding and Powell (1987)	M	M	M	M
IT13887	Ritchie and Edwards (1996)	H	H	M	H
IT16660	Schmid and Telaro (1990)	M	M	M	M
IT16484	Shayer and Beasley (1987)	M	M	M	M
IT13507	Strang and Shayer (1993)	M	M	M	M
IT15898	Tenenbaum (1986)	H	H	H	H
IT16670	Tzuriel and Alfassi (1994)	M	M	M	M
IT13466	Ward and Traweek (1993)	M	M	M	M

4.2 Effect sizes and publication bias

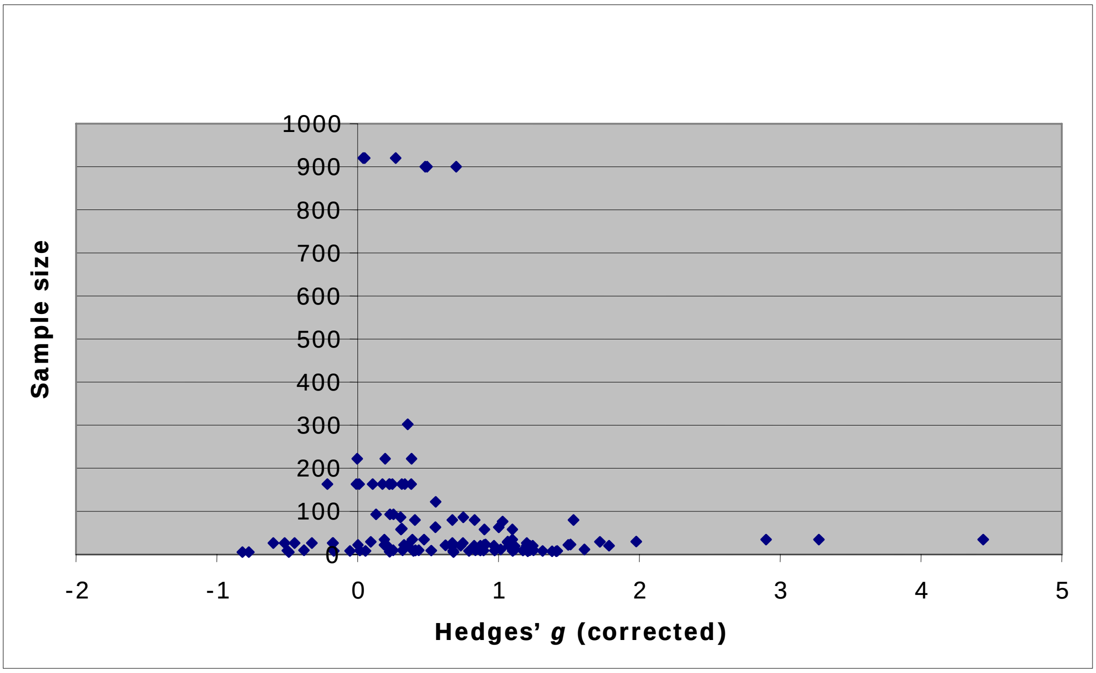
4.2.1 Effect sizes

The majority of studies (19) reported pre- and post-test means and standard deviations. For the ten remaining studies, the main effect size was calculated from *t*-tests (6) and *f*-values (4), using the meta-analysis tools within EPPI-Reviewer.

4.2.2 Publication bias

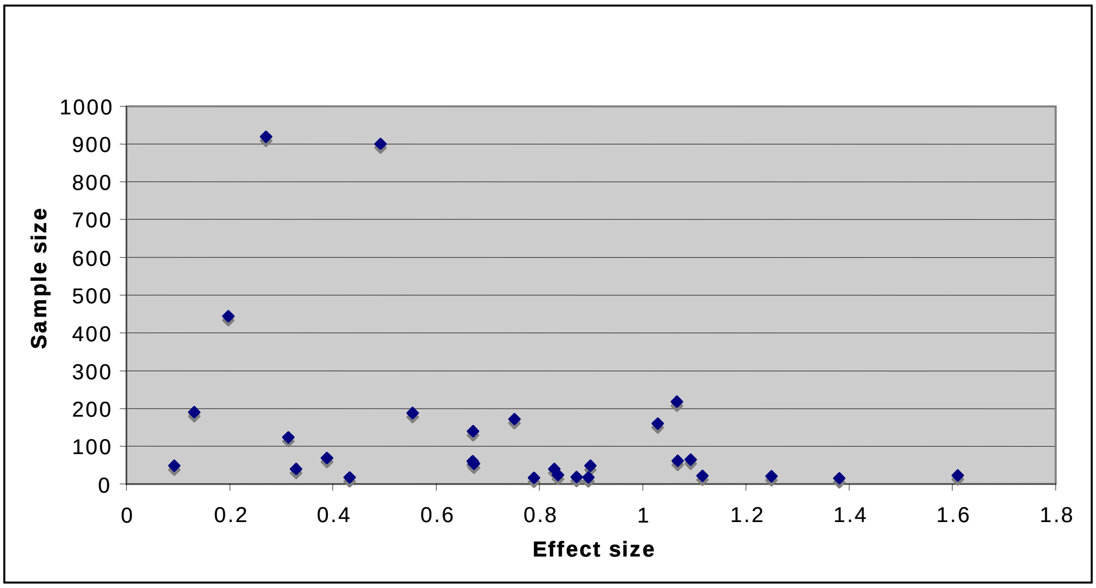
We investigated the relationship between the sample sizes and the effects of the interventions using a ‘funnel’ plot (Lipsey and Wilson, 2001). If there is no publication bias, then the included studies should form an inverted ‘funnel’ shape, with the studies with the largest samples at the top.

Figure 4.1: Funnel plot of all effects



However, a slightly different picture emerges when one key outcome is selected per study. This key outcome was identified by making a judgement about the most appropriate measure according to the aims of the study. Where the intention was to investigate the impact of a thinking skills approach on pupils' thinking or reasoning the main cognitive measure (such as Raven's Progressive Matrices), was selected; where the intention was to improve pupils' performance in a specific curriculum subject (such as reading comprehension) the most appropriate curriculum test measure was selected.

Figure 4.2: Funnel plot of sample size and main effect



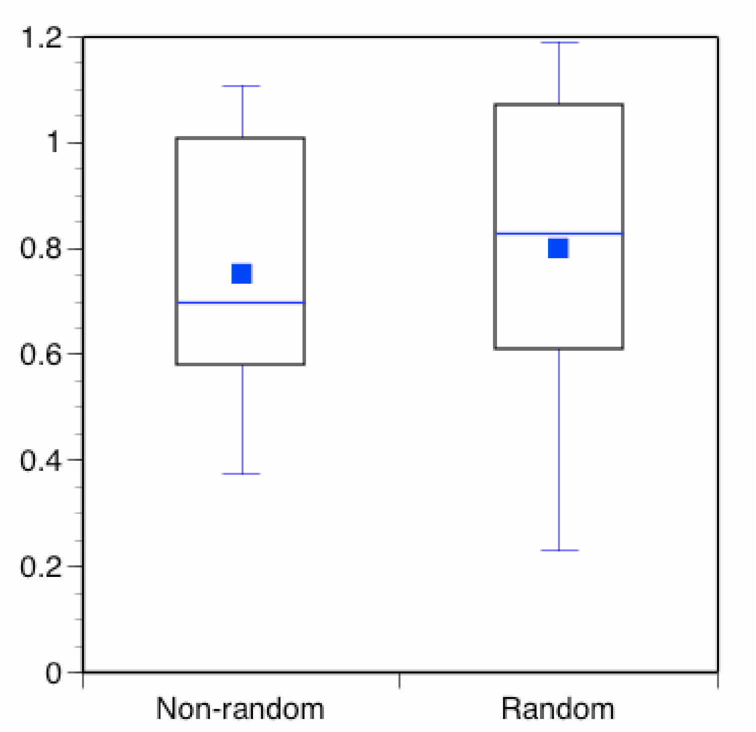
This funnel plot suggests that there may be some publication bias as the largest studies have lower effects. It may also indicate that small scale interventions are more effective. However, the small number of large studies (and the relatively small number of studies overall) means that it is difficult to draw firm conclusions. One of the key statistical assumptions in this technique is that the larger studies more closely represent the true estimate of effect as it is assumed that they are less subject to variability (Greenhouse and Iyengar, 1994). However, this assumption may be questioned as fidelity of implementation, for example, may be affected by the scale of an intervention which may itself be an important variable (through the quality of training of teachers to undertake the intervention and their individual beliefs and commitment to a particular programme or approach).

4.3 Study heterogeneity

One of the issues with a quantitative synthesis is how comparable the various studies included in the review are (Kulik and Kulik, 1989b). The studies selected all met our criteria for inclusion in the study and came under our agreed criteria for thinking skills interventions (which the members of the Review Group were able to apply to studies with a high degree of reliability). However, the studies are of different thinking skills programmes and approaches, used in different educational settings, across different countries, subjects and age groups. As a result, a number of analyses were undertaken to look at the range and distribution of effects.

We looked at the different study designs to see whether there was any difference between studies where pupils were randomly allocated to intervention and control groups and where the allocation was not done randomly (such as when using existing class groupings). A box and whisker plot was used to compare the two groups (Figure 4.3). This shows the distribution and range of impact of random and non-randomly assigned interventions. The ‘box’ represents the quartile values of the data (i.e. 50%) and the whiskers the extent of the range (top quartile and bottom quartile). The median is shown by the line towards the middle of the box and the mean by the small shaded square.

Figure 4.3: Box plot of effect size by allocation of pupils in the study



The relationship between study quality and effect size investigated further to see if there was any association (Table 4.2). This is a contested issue in both education meta-analysis and other fields of research (e.g. MacLehose *et al.*, 2000) where there is often an assumption that the quality of the study and its reporting will be linked to the effect size and that high quality studies will provide a more accurate picture or a truer estimate of effect and further an implication that low quality studies will report larger effect sizes.

Table 4.2: Weight of evidence and effect size

Weight of evidence[9]	Effect size[10]	Confidence interval	Number of studies
High	0.57	0.41, 0.73	15
Medium	0.77	0.49, 1.06	11
Low	0.43	0.11, 0.75	3

Our findings are similar to those reported by Marzano (1998, p 86) where, with a sample of 1,370 studies, he found no significant relationship between the methodological quality of the studies and the effect sizes within a given study. These findings are also consistent with those reported by Smith *et al.* (1980) and Lipsey and Wilson (1993) who found that studies rated high on methodological quality had an effect size of 0.40, while studies rated low on methodological quality had an effect size of 0.37.

[9] Weight of evidence A or study quality: see Table 4.1.
[10] Random effects model; one key outcome per study selected according to study aims.

4.4 Quantitative synthesis of evidence

These studies contained a large number of effect sizes using a wide range of measures (122 effects reported in, or calculated from, data in the 29 studies). A crude analysis was undertaken to identify the broad overall effect to see how interventions characterised as 'thinking skills' compared with other researched interventions. This approach, however, may give a misleading overall result as the effect sizes are not independent of each other (as some effects come from the same study) so three further analyses were undertaken: first, with one key outcome identified for each study (the main cognitive outcome); second, with one main curriculum outcome identified (where available) per study; and third, with one main affective outcome identified per study.

Even taking this approach, tests for heterogeneity were all significant. The studies had a range of interventions and different results; it is difficult to discover the precise reasons for this heterogeneity. As a result, we used the random effects model for quantitative synthesis as this approach places less emphasis on the larger studies than the fixed effects model and is therefore more appropriate for combining heterogeneous groups of studies. Analysis of these studies indicate that thinking skills approaches are effective in improving pupils' learning and that they have a positive effect on pupils' attitudes or beliefs. The quantitative synthesis of this impact found an effect size of 0.62 for the main cognitive measure (such as tests of reasoning or non-verbal measures, such as Raven's Progressive Matrices) and an effect size of 0.62 for the main curriculum outcomes (such as reading, mathematics or science tests). These effect sizes indicate that an 'average' class of pupils who received such interventions would move from 50th place in a rank of 100 similar classes to about 26th (a gain of 24 percentile points). A table of all of the effects can be found in Appendix 4.1. A summary of the effect sizes and types of outcomes is presented in Table 4.3[11].

Table 4.3: Overall effect sizes and outcome measures

Type of outcome	ES	CI	Number of effects / Number of studies
All outcomes	0.74	0.63, 0.85	122 effects from 29 studies[12]
Cognitive outcomes	0.62	0.54, 0.69	29 effects from 29 studies[13]
Curricular outcomes	0.62	0.45, 0.80	19 effects from 19 studies[14]
Affective outcomes	1.44	0.47, 2.41	6 effects from 6 studies[15]

However, some caution is required in interpreting this meta-analysis as there are considerable differences in the thinking skills approaches and programmes and included in the analysis. Forest plots for the impact on cognitive, curricular and affective outcomes are presented in Figures 4.4–4.6 to show the range of impact across these groupings.

[11] Study heterogeneity was calculated using the DerSimonian and Laird (1986) random effects method by the meta-analysis software in the EPPI-Reviewer database together with overall effect sizes.
[12] Heterogeneity statistic $Q = 1.18E+03$, $df = 125$, $p = 0$. Test statistic (combined effect) $z = 13.4$ $p < 0.001$
[13] $Q = 86.3$, $df = 27$, $p = 4.1E-08$, $z = 9.05$ $p < 0.001$
[14] $Q = 72.8$, $df = 26$ $p = 2.6E-06$, $z = 16.8$, $p < 0.001$
[15] $Q = 87.6$, $df = 5$, $p = 0$, $z = 2.92$, $p = 0.00346$

Figure 4.4: Forest plot of main cognitive outcome per study

Item	Effect (CI)	Weight %	Size
Main outcome - cog			
Adey and Shayer (1990) Acceleratingthe developmentof formal thinking	0.09(-0.49, 0.67)	3.4	48
Adey, Robertson & Venville(2002) Effects ofcognitive acceleration	0.56(0.25, 0.86)	5.7	188
Cardelle-Elawar (1992) Effects of Teaching Metacognitive Skills	1.09(0.56, 1.63)	3.7	62
Chang & Barufaldi (1999) The use of a problem-solving model	0.75(0.44, 1.06)	5.7	172
Collings (1994) Some fundamental questions about scientific thinking	0.67(0.11, 1.23)	3.6	60
Csapo (1992) Improving Operational Abilities in Children	0.70(0.56, 0.83)	7.1	900
De Koning & Hamers (1999) Teaching inductive reasoning	0.84(-0.02, 1.69)	2.1	24
Greenberg (2000) Attending to hidden needs	0.90(0.30, 1.50)	3.3	48
Haywood et al. (1988) Cognitive education with deaf adolescents	0.67(0.12, 1.23)	3.6	53
Hoek et al (1999) The effects of social and cognitive strategy on math	0.20(0.01, 0.38)	6.8	444
Iqbal & Shayer (2000) Acceleratingthe developmentof formal thinking	1.07(0.78, 1.35)	5.9	218
Kaniel & Reichenberg (1992) Instrumental Enrichment - effects	0.67(0.33, 1.02)	5.3	140
Kramaski & Mevarech (1997) Cognitive-metacognitive training	0.39(-0.09, 0.87)	4.2	68
Maqsud (1998) Effects of metacognitive instruction on mathematics	0.83(0.18, 1.48)	3.0	40
Martin (1984) Cognitive modification	0.87(-0.11, 1.85)	1.7	18
Mercer et al. (1999) Children's talk and reasoning	0.31(-0.04, 0.67)	5.2	124
Muttart (1984) Effects of Instrumental Enrichment	0.89(-0.12, 1.91)	1.6	17
Oladunni (1988) The effectiveness of problem solving	1.03(0.70, 1.36)	5.5	161
Riding & Powell (1987) Effect of critical thinking activities	0.79(-0.24, 1.82)	1.6	16
Ritchie & Edwards (1996) Creative thinking instruction	0.33(-0.30, 0.96)	3.1	40
Schmid & Telaro (1990) Concept mapping as an instructional strategy	0.43(-0.55, 1.41)	1.7	17
Severino (1993) Developing creative thinking	1.38(0.21, 2.55)	1.3	15
Shayer & Beasley (1987) Does Instrumental Enrichment Work?	1.25(0.27, 2.23)	1.7	20
Strang & Shayer (1993) Enhancing achievement through thinking skills	1.12(0.18, 2.05)	1.8	21
Tenenbaum (1986) The effect of quality of instruction	1.07(0.53, 1.61)	3.7	61
Tzuriel & Alfassi (1994) Cognitive and motivational modifiability	0.13(-0.15, 0.42)	5.9	191
Ward & Traweek (1993) Application of a metacognitive strategy	1.61(0.65, 2.58)	1.7	23
	0.68(0.54, 0.83)		

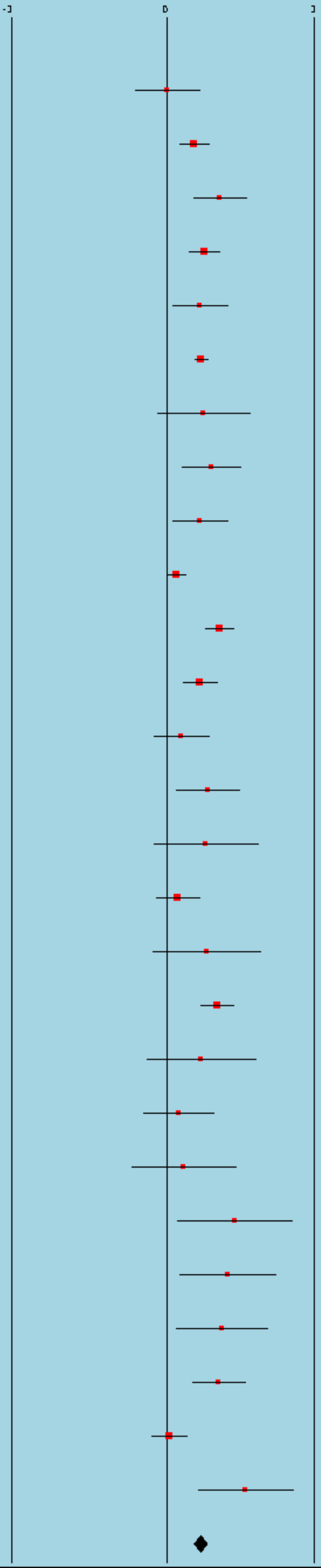


Figure 4.5: Forest plot of main curriculum outcome per study

Item	Effect (CI)	Weight %	Size
Main outcome - curric			
Adey, Robertson & Venville (2002) Effects of cognitive acceleration	0.36(0.17, 0.55)	9.5	468
Cardelle-Elawar (1992) Effects of Teaching Metacognitive Skills	0.74(0.23, 1.26)	5.6	62
Chang & Barufaldi (1999) The use of a problem-solving model	0.31(0.00, 0.61)	8.2	172
Collings (1994) Some fundamental questions about scientific thinking	1.12(0.53, 1.70)	5.0	60
De Koning & Hamers (1999) Teaching inductive reasoning	1.22(0.33, 2.12)	2.9	24
Greenberg (2000) Attending to hidden needs	0.90(0.49, 1.31)	6.8	102
Haywood et al. (1988) Cognitive education with deaf adolescents	-0.17(-0.71, 0.37)	5.4	53
Hoek et al (1999) The effects of social and cognitive strategy on math	0.38(0.20, 0.57)	9.5	444
Kramaski & Mevarech (1997) Cognitive-metacognitive training	0.39(-0.09, 0.87)	6.0	68
Maqsud (1998) Effects of metacognitive instruction on mathematics	0.87(0.22, 1.52)	4.4	40
Martin (1984) Cognitive modification	0.52(-0.42, 1.47)	2.6	18
Muttart (1984) Effects of Instrumental Enrichment	0.97(-0.05, 2.00)	2.3	17
Oladunni (1988) The effectiveness of problem solving	1.03(0.70, 1.36)	7.8	161
Schmid & Telaro (1990) Concept mapping as an instructional strategy	0.43(-0.55, 1.41)	2.5	17
Shayer & Beasley (1987) Does Instrumental Enrichment Work?	0.32(-0.57, 1.20)	2.9	20
Strang & Shayer (1993) Enhancing achievement through thinking skills	1.12(0.18, 2.05)	2.7	21
Tenenbaum (1986) The effect of quality of instruction	1.06(0.51, 1.62)	5.2	58
Tzuriel & Alfassi (1994) Cognitive and motivational modifiability	0.23(-0.05, 0.52)	8.4	191
Ward & Traweek (1993) Application of a metacognitive strategy	1.61(0.65, 2.58)	2.6	23
	0.62(0.45, 0.80)		

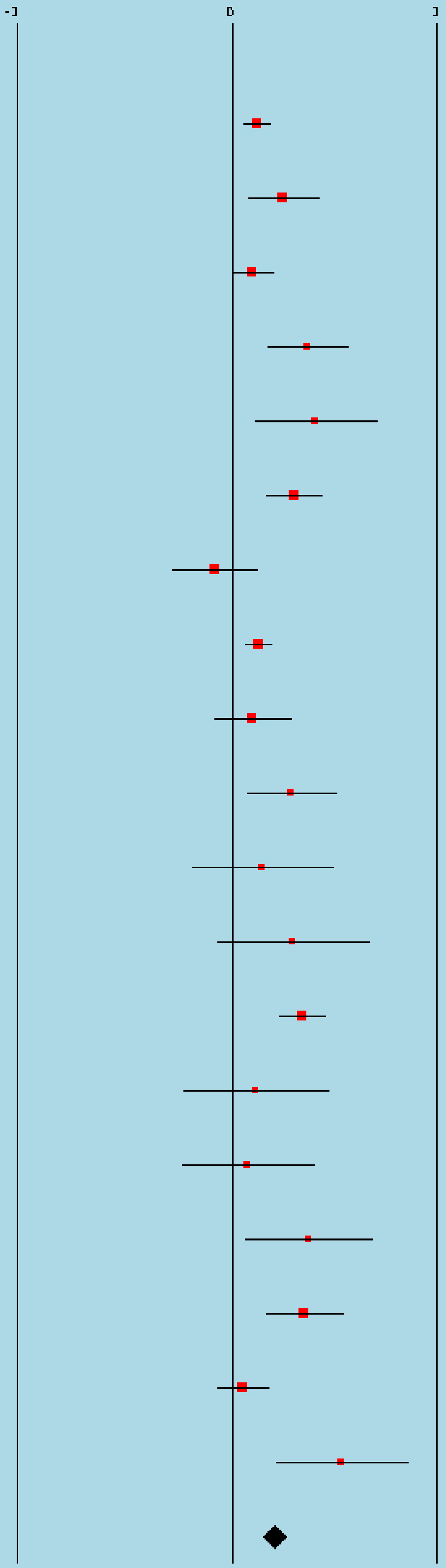
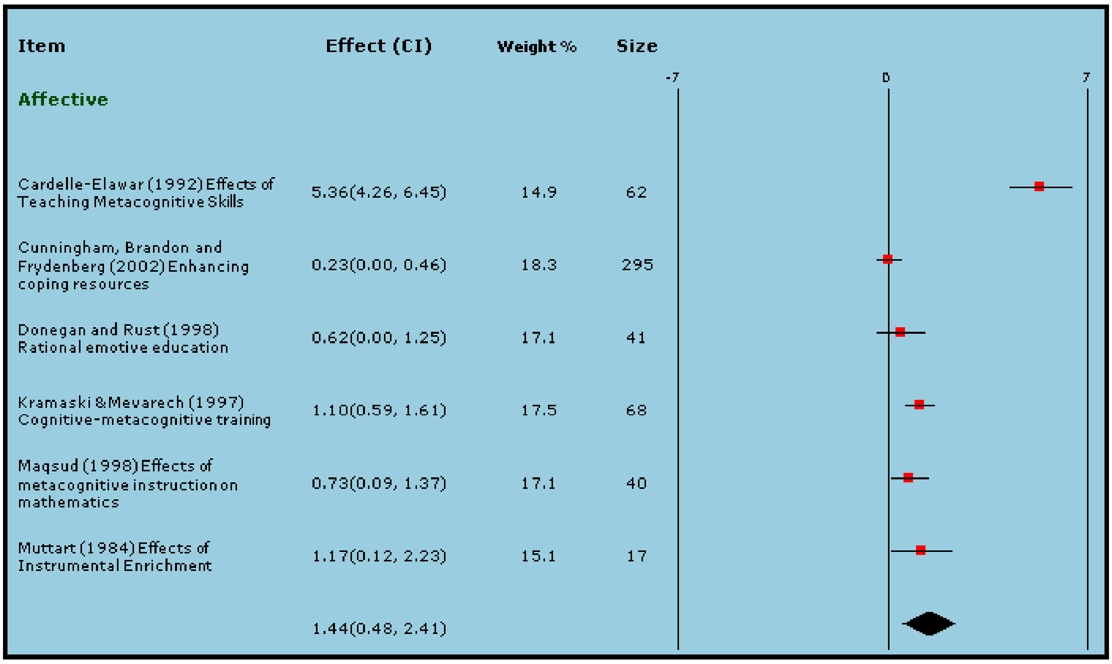


Figure 4.6: Forest plot of main affective outcome per study



4.4.1 Impact of thinking skills programmes and approaches

A further analysis was undertaken according to the kind of thinking skills intervention.

Table 4.4: Effect sizes for different programmes

Intervention	Effect size[16]	Confidence interval	Number of studies
Instrumental enrichment	0.58	0.28, 0.88	7
Cognitive acceleration[17]	0.61	0.25, 0.96	4
Metacognitive strategies[18]	0.96	0.76, 1.16	5

Feuerstein's instrumental enrichment is one of the most extensively researched of the thinking skills programmes since its development in the 1970s. Our results broadly concur with those of Romney and Samuels (2001), whose meta-analysis found moderate overall effects and an effect size of 0.43 on reasoning ability (p 28). Our findings were of the same order with an overall effect size of 0.58 (one main effect from seven studies) and an effect size of 0.52 on tests of reasoning (one main effect from four studies).

Our first review concluded, on the basis of the narrative synthesis undertaken, that there was evidence that the cognitive acceleration family of interventions

[16] Random effects model: one key cognitive outcome per study selected according to study aims

[17] A case can be made that the impact of cognitive acceleration programmes increases over time (Adey and Shayer, 1993); this analysis uses the immediate post-test scores for comparability.

[18] Although this is a broad category and a case can be made that almost all the thinking skills approaches involve metacognitive elements, we identified interventions which explicitly focused on metacognitive skills and strategies so as to make a comparison with Marzano (1998), who identified metacognitive approaches as likely to have greater impact.

seemed to have clear benefits (Higgins *et al.*, 2004, p 40). Quantitative synthesis appears to validate this conclusion with an effect size of 0.61 resulting from a synthesis of they key outcome from four studies[19].

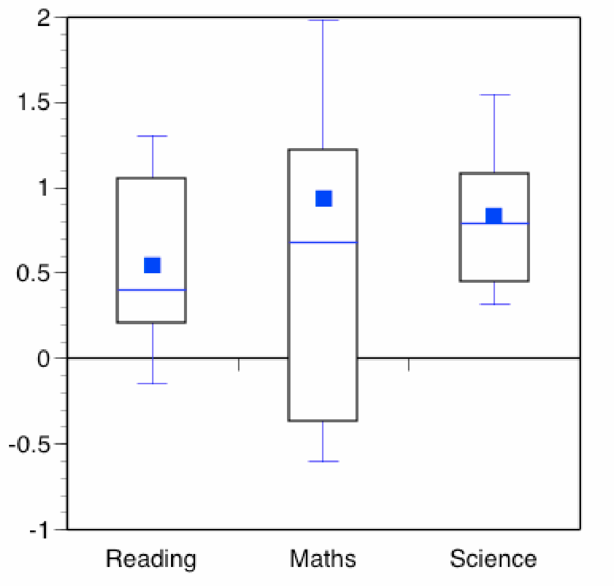
Marzano's (1998) theory-driven meta-analysis identified the overall effect size for instructional techniques that focused on the metacognitive system as 0.72 (p 88). We identified five studies which explicitly identified metacognitive strategies as the focus for the intervention. These studies had an overall effect size of 0.96, based on one key outcome identified per study. This again appears to validate one of the conclusions of our narrative synthesis regarding the value of metacognitive approaches.

4.4.2 Impact on curriculum outcomes

Table 4.5: Effect sizes for different curriculum areas

Subject	Effect size[20]	Confidence Interval	Number of studies
Reading[21]	0.48	0.24, 0.71	7
Mathematics	0.89	0.50, 1.29	9
Science	0.78	0.47, 1.09	5

Figure 4.7: Box plot of curriculum outcomes



We also investigated the impact on the three main curriculum areas tested in the studies included in the review. There appears to be greater impact on pupils'

[19] It should also be noted that the effect sizes used to calculate this are based on immediate post-tests and there is evidence from CASE that such effects may increase over time (Adey and Shayer, 1993).
[20] Random effects model; one key outcome selected according to study aims.
[21] Reading (and where available reading comprehension) was selected as the key outcome as not all of the studies were in English speaking schools.

attainment in Mathematics (0.89) and Science (0.78) than on Reading (0.4). This may be due to the specificity of the skills and content taught in the thinking skills programmes and approaches which focus on mathematics and science.

4.5 In-depth review: quality-assurance results

All studies were data-extracted by at least two reviewers independently. Seven of the studies were data-extracted by members of the EPPI-Centre. All differences in coding were resolved and an agreed version of the data-extraction uploaded prior to the synthesis stage of the review.

4.6 Nature of actual involvement of users in the review and its impact

Users were fully integrated into the core review team and took a full and active part in each stage of the review. However, the online data extraction was largely completed by the members of the team from higher education institutions (HEAs) and then discussed with the wider group. This was a direct consequence of the time pressures and the lack of funding to release school-based staff for the length of time required to complete the exercise. We regard this as an issue of the best deployment of resources rather than any invocation of a hierarchy of expertise. Each member of the core review team was involved in a paper exercise of data extraction so that they were equally well informed as to the nature of the process. The subsequent synthesis of evidence was discussed at team meetings and its significance for our review question validated via consultation with colleagues from across the Review Group.

5. FINDINGS AND IMPLICATIONS

In this review, we have reported on the results of a systematic search and quantitative synthesis or meta-analysis of studies which have evaluated the impact of the use of thinking skills approaches in schools. The identified studies contained evidence of the impact of such programmes and approaches on pupils' attainment and attitudes compared with a control or comparison group.

The quantitative synthesis indicates that thinking skills programmes and approaches are effective in improving the performance on cognitive measures (such as Raven's progressive matrices) with an overall effect size of 0.62. These approaches also have a considerable impact on curricular outcomes with the same effect size of 0.62. However, there is considerable variation in the impact of such approaches and caution is needed in interpreting this overall figure. The studies included in the review describe the impact of a range of interventions across a range of different educational contexts.

5.1 Summary of principal findings

5.1.1 Identification of studies

Twenty-six of the studies identified for this meta-analysis were obtained from a database which resulted from the first thinking skills review (Higgins et al., 2004); a further three resulted from updating the original search and applying the more stringent criteria required for a quantitative synthesis.

5.1.2 Nature of studies selected for meta-analysis

Twenty-nine studies were identified which contained quantitative data on pupils' attainment and attitudes suitable for meta-analysis. The studies come from a range of countries around the world, with half set in the US and UK. The studies broadly cover the ages of compulsory schooling (5–16) and included studies set in both primary and secondary schools. A number of named thinking skills interventions are included, such as Feuerstein's instrumental enrichment (FIE) and cognitive acceleration through science education (CASE) as well as studies which report a more general thinking skills approach (such as the development of metacognitive strategies).

5.1.3 Synthesis of findings from studies in the in-depth review

Overall, the quantitative synthesis indicates that, when thinking skills programmes and approaches are used in schools, they are effective in improving pupils' performance on a range of tested outcomes (relative to those who did not receive thinking skills interventions). The magnitude of the gains found appears to be important when compared with the reported effect sizes of other educational interventions.

Meta-analysis does appear to be finding broadly consistent messages in the educational research literature. Our study found an overall mean effect of 0.62 for the main effect of the studies, larger than the mean effect found by Sipe and Curlette (1997) of 0.375, larger than Hattie's database of meta-analyses at 0.4 (Hattie, 1999) but very similar to the overall figure for interventions across the knowledge, cognitive, meta-cognitive and self-system domains reported by Marzano (1998, p 76) of 0.65. In particular, our study identified metacognitive interventions as having relatively greater impact (0.96), as did Marzano.

Looking at a smaller part of our review, Feuerstein's instrumental enrichment is one of the most extensively researched of the thinking skills programmes, Our results broadly concur with those of Romney and Samuels (2001), whose meta-analysis found moderate overall effects and an effect size of 0.43 on reasoning ability (p 28). Our findings were of the same order with an overall effect size of 0.58 (one main effects from each of seven studies included) and an effect size of 0.52 on tests of reasoning (one main effect from four studies).

There is some indication that the impact of thinking skills programmes and approaches may vary according to subject. In our analysis, there was relatively greater impact in mathematics and science compared with reading.

Overall, this review suggest to us that the findings from meta-analysis are worth considering as a part of the story of 'what works' in education by offering comparative information about *how well* different interventions work. This echoes Hattie's (1999) plea that:

- 'We need to make *relative* statements about what impacts on student work.
- We need estimates of magnitude as well as statistical significance – it is not good enough to say that this works because lots of people use it, but that this works because of the magnitude of impact.
- We need to be building a model based on these relative magnitudes of effects.'

5.2 Strengths and limitations of this systematic review

Of course interpreting the results of a quantitative synthesis and comparing the relative magnitude of effects will always be challenging. The messages may well not be consistent across different contexts. This in turn suggests that we need better descriptive reporting of what happens in educational interventions in terms of the processes of teaching and learning which will support clearer conceptual analysis of ways that we can identify similarities and differences across those different contexts. The term 'thinking skills' is clearly a loose construct. Although we (and the teachers that we worked with) were able to use it to categorise interventions reliably (Higgins *et al*., 2004) it is a very broad term and the programmes and approaches under this umbrella can look very different in the classroom.

Possible weaknesses can also be pointed out from a methodological perspective, as Lipsey and Wilson (2001) indicate, 'Meta-analysis results are only as good as the studies that are included in the meta-analysis' (p 157). Critics of the experimental and quasi-experimental research approaches in education or those wishing to identify definitive causal links among the host of variables in educational research may well find much to criticise in meta-analysis, but pragmatists or those who adopt a more positivist approach can perhaps use it as a useful tool to impose some order on the apparently diverse and contradictory findings from educational research. It is certainly a valuable tool to organise and analyse large amounts of data. What is more challenging is to interpret the outcomes of such analyses, especially for different audiences. The use of the term 'thinking skills' in relation to the educational research literature has identified a collection of research studies which have an above average impact on learning outcomes. This suggests that teachers' interest and enthusiasm for such approaches is well founded (Baumfield and Oberski, 1998; Leat and Higgins, 2002) as such approaches tend to have positive effects, over and above what you would usually expect for an educational intervention.

Our interpretation would be that it is possible to make a 'fuzzy generalisation' (Bassey, 2000) about thinking skills approaches and that the use of meta-analysis is one possible approach to use it as a 'best estimate of trustworthiness' (BET). The implications are perhaps clearer at policy level that the use of thinking skills programmes and approaches should be encouraged. There are already signs that the Department for Education and Skills has effectively acted on this evidence in the materials for the National Key Stage 3 Strategy and in the development of a database of thinking skills resources for primary schools on the Standards Site[22].

For schools and teachers, the specific implications for classroom practice are less clear. There is certainly mounting evidence that adopting approaches which make aspects of thinking explicit or which focus on particular kinds of thinking are successful at raising attainment – particularly metacognitive approaches (Marzano, 1998) – nor cognitively demanding interventions, such as problem-solving and hypothesis-testing, or those that enhance surface and deep learning (Hattie, 2004)). The meta-analysis described in this review adds weight to this growing body of evidence. However, the variation in the impact of thinking skills approaches by age or gender (for example, Adey and Shayer, 1993; Csapó, 1992) combined with the differences between the programmes themselves make it difficult to offer more than Bassey's BET without further research and analysis.

Strengths

- *A strength of the review is the extent of the literature included in the review process, building on the mapping of this literature and narrative synthesis completed for the first review.* This systematic review is based on an extensive search across a wide range of literature and descriptive map of studies included which helps to contextualise the research which has emerged through the process for in-depth analysis.

- *A strength of the review is the use of meta-analysis to provide a quantitative synthesis of the research literature and an overall estimate of the impact of such approaches*. Although the use of meta-analysis is controversial, it provides a way to aggregate the research literature which complements narrative reviews.

- *A strength of the review is the quantitative synthesis which contextualises thinking skills research within the broader field of educational research*. This

[22] http://www.standards.dfes.gov.uk/thinkingskills/

suggests that thinking skills programmes and approaches tend to be more successful at improving pupils' performance on both measures of cognitive outcomes and curriculum tests.

- *A strength of this review is the close involvement of users groups in setting and refining the questions, and interpreting and disseminating the findings*. The Review Group was keen to ensure that the perspectives of practitioners were included in the review and involved those working in and with experience of school at every stage, in order to maintain the link between research, the interpretation of that research, and the development of practice in schools.

Limitations

The Review Group was conscious throughout of the complexity of the question they were trying to answer and limitations of the data in the separate and diverse studies examined for the review. Identifying the causes of impact on pupil attainment in classrooms is complex and the focus of the studies did not always acknowledge this complexity. In particular, the Review Group found the following:

- There was a tendency for studies to report little about the programmes themselves or aspects of their implementation and use in classrooms (such as changes in teaching and learning processes). It is therefore difficult to draw conclusions about any common features of programmes and approaches which may account for the positive impact reported.

- Within the timescale and resources for the review, 29 studies were identified and synthesised within the timescale and resources for the review. A larger number of studies would enable further analysis (such as by age or subject) to make more specific recommendations for practitioners and policy makers.

- The review used a broad definition of thinking skills for its focus. As a result there was considerable statistical heterogeneity in the results of the studies which indicates that caution is required in combining the effects and interpreting the findings.

- Meta-analysis itself is subject to a number of limitations and criticisms, leaving this review open to the same critiques.

5.3 Implications

Thinking skills programmes have been extensively used around the world for a number of years. There is a growing body of accumulated evidence that they are effective at improving pupils' performance on cognitive and curriculum tests when they are researched in school settings. Their effect is relatively greater than most other researched educational interventions. This review strengthens this evidence-base.

For practitioners, thinking skills programmes and approaches are likely to improve pupils' learning. Their use in schools should therefore be supported. Some caution is required as there is some variation in the impact of such approaches according to subject, age and gender. This suggests that their use needs to be matched to the particular teaching context and monitored critically to ensure potential benefits.

For policy-makers, thinking skills programmes and approaches are an effective way to improve teaching and learning, and their use in schools should be encouraged. However, as it is not clear to what extent the benefits are due to specific aspects of the programmes and their implementation or the changes in teaching and learning which ensue, it is not possible to provide precise recommendations.

Further research is needed to clarify the particular causes of the benefits and where thinking skills programmes and approaches have most impact (such as on different age groups or in different areas of the curriculum). In particular, the impact of thinking skills programmes and approaches on teaching and learning processes needs to be related to improvements in outcomes to identify the means by which the impact occurs.

Researchers and journal editors should note that studies were often excluded because basic information relating (such as number of pupils involved) was not included in the published papers. Moreover, the details of sampling strategies and full set of results were frequently omitted. Abstracts sometimes referred to data which was not then reported in detail. Journals which used structured abstracts were more likely to contain more accurate and more complete information to support the systematic reviewing process.

6. REFERENCES

6.1 Studies included in map and synthesis

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

Adey P, Robertson A, Venville G (2002) Effects of a cognitive acceleration programme on Year 1 pupils. *British Journal of Educational Psychology* **72:** 1–25.

Cardelle-Elawar M (1992) Effects of teaching metacognitive skills to students with low mathematics ability. *Teaching and Teacher Education* **8:** 109–121.

Chang C-Y, Barufaldi JP (1999) The use of a problem-solving-based instructional model in initiating change in students' achievement and alternative frameworks. *International Journal of Science Education* **21:** 373–388.

Collings JN (1994) Some fundamental questions about scientific thinking. *Research in Science and Technological Education* **12:** 161–173.

Csapó B (1992) Improving operational abilities in children. In: Demetriou A, Shayer M, Efklides A (eds) *Neo-Piagetian Theories of Cognitive Development.* London: Routledge, pages 144–159.

Cunningham EG, Brandon CM, Frydenberg E (2002) Enhancing coping resources in early adolescence through a school-based program teaching optimistic thinking skills. *Anxiety Coping and Stress* **15**: 369–381.

De Koning E, Hamers JHM (1999) Teaching inductive reasoning: theoretical background and educational implications. In: Hamers JHM, Van Luit JEH, Csapó B (eds) *Teaching and Learning Thinking Skills.* Abingdon: Swets and Zeitlinger, pages 157–188.

Donegan AL, Rust JO (1998) Rational emotive education for improving self concept in second grade students. *Journal of Humanistic Education and Development* **36:** 248–256.

Greenberg KH (2000) Attending to hidden needs: the cognitive enrichment advantage perspective. *Educational and Child Psychology. Special Issue: Psychological influences upon educational intervention* **17:** 51–69.

Haywood HC, Towery-Woolsey J, Arbitman-Smith R, Aldridge AH (1988) Cognitive education with deaf adolescents: effects of instrumental enrichment. *Topics in Language Disorders* **8:** 23–40.

Hoek D, Van den Eeden P, Terwel J (1999) The effects of integrated social and cognitive strategy instruction on the mathematics achievement in secondary education. *Learning and Instruction* **9**: 427–448.

Iqbal HM Shayer M (2000) Accelerating the development of formal thinking in Pakistan secondary school students: achievement effects and professional development issues. *Journal of Research in Science Teaching* **37:** 259–274.

Kaniel S, Reichenberg R (1992) Instrumental enrichment: effects of generalization and durability with talented adolescents. *Gifted Education International* **8:** 128–135.

Kramarski B, Mevarech ZR (1997) Cognitive-metacognitive training within a problem-solving based Logo environment. *British Journal of Educational Psychology* **67:** 425–445.

Maqsud M (1998) Effects of metacognitive instruction on mathematics achievement and attitude towards mathematics of low mathematics achievers. *Educational Research* **40:** 237–243.

Martin DS (1984) Cognitive modification for the hearing impaired adolescent: the promise. *Exceptional Children* **51:** 235–242.

Mercer N, Wegerif R, Dawes L (1999) Children’s talk and the development of reasoning in the classroom. *British Educational Research Journal* **25:** 95–111.

Muttart K (1984) Assessment of effects of instrumental enrichment cognitive training. *Special Education in Canada* **58:** 106–108.

Naval-Severino T (1993) Developing creative thinking among intellectually able Filipino children from disadvantaged urban communities. *Gifted Education International* **9:** 119–123.

Oladunni MO (1998) An experimental study on the effectiveness of metacognitive and heuristic problem solving techniques on computational performance of students in mathematics. *International Journal of Mathematical Education in Science and Technology* **29:** 867–874.

Riding RJ, Powell SD (1987) The effect on reasoning, reading and number performance of computer presented critical thinking activities in five-year-old children. *Educational Psychology* **7:** 55–65.

Ritchie SM, Edwards J (1996) Creative thinking instruction for aboriginal children. *Learning and Instruction* **6:** 59–75.

Schmid RF, Telaro G (1990) Concept mapping as an instructional strategy for high school biology. *Journal of Educational Research* **84:** 78–85.

Shayer M, Beasley F (1987) Does instrumental enrichment work? *British Educational Research Journal* **13:** 101–119.

Strang J, Shayer M (1993) Enhancing high school students’ achievement in chemistry through a thinking skills approach. *International Journal of Science Education* **15:** 319–337.

Tenenbaum G (1986) The effect of quality of instruction on higher and lower mental processes and on the prediction of summative achievement. *Journal of Educational Research* **80:** 105–114.

Tzuriel D, Alfassi M (1994) Cognitive and motivational modifiability as a function of the Instrumental Enrichment (IE) Program. *Special Services in the Schools* **8:** 91–128.

Ward L, Traweek D (1993) Application of a metacognitive strategy to assessment, intervention and consultation: a think-aloud technique. *Journal of School Psychology* **31:** 469–485.

6.2 Other references used in the text of the report

Adey P, Shayer M (1993) An exploration of long-term far-transfer effects following an extended intervention program in the high school science curriculum. *Cognition and Instruction* **11:** 1–29.

Adey P, Shayer M (1994) *Really Raising Standards.* London: Routledge.

Adey PS, Shayer M, Yates C (1995) *Thinking Science: The Curriculum Materials of the CASE Project* London: Thomas Nelson and Sons.

Bassey M (2000) Reviews of Educational Research: Annex II: 'BETS on Homework'. *Research Intelligence* **71**: 28–29.

Baumfield VM, Oberski IO (1998) What do teachers think about thinking skills? *Quality Assurance in Education* **6**: 44–51.

Black P, Wiliam D (1998) Assessment and classroom learning. *Assessment in Education* **5**: 7–71.

Blagg N, Ballinger M, Gardner R (1988) *Somerset Thinking Skills Course Handbook*. Oxford: Basil Blackwell.

Bloom B, Englehart M, Furst E, Hill W, Krathwohl D (1956) *Taxonomy of educational objectives: the classification of educational goals. Handbook I: Cognitive domain.* New York, Toronto: Longmans, Green.

Brown AL, Palincsar AS (1989) Guided co-operative learning and individual knowledge acquisition. In: Resnick L (ed) *Knowing, Learning, and Instruction: Essays in Honor of Robert Glaser*. Hillsdale, NJ: Erlbaum, pages 393–451.

Chambers EA (2004) An introduction to meta-analysis with articles from the Journal of Educational Research (1992–2002). *Journal of Educational Research* **98**: 35–44.

Cohen J (1988) *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

Cooper H, Dorr N, Bettencourt BA (1995). Putting to rest some old notions about social science. *American Psychologist* **50**: 111–112.

Cotton K (1991) Teaching Thinking Skills. School Improvement Research Series (SIRS). (On-line serial from NW Regional Educational Laboratory). Available at http://www.nwrel.org/scpd/sirs/6/cu11.html

Csapó B (1992) Improving operational abilities in children. In: Demetriou A, Shayer M, Efklides A (eds) *Neo-Piagetian Theories of Cognitive Development: Implications and Applications for Education.* London: Routledge, pages 144–159.

Deeks JJ, Altman DG, Bradburn MJ (2001) Statistical methods for examining heterogeneity and combining results from several studies in meta-analysis. In: Egger M, Davey Smith G, Altman D (eds) *Systematic Reviews in Health Care: Meta-Analysis in Context (2nd edition)*. London: BMJ Books, pages 285–312.

DerSimonian R, Laird N (1986) Meta-analysis in clinical trials. *Controlled Clinical Trials*. **7**: 177–188.

DfES (2004a) *Teaching and Learning in the Foundation Subjects*. DfES 09-2004 DfES 0123-2004 G. Annesley, Nottingham: DfES Publications.

DfES (2004b) *Excellence and Enjoyment, Learning and Teaching in the Primary Years*. DfES 09-2004 DfES 0518-2004 G. Annesley, Nottingham: DfES Publications.

DfES (2005) *Leading in Learning: Developing Thinking Skills at Key Stage 3*. 01-2005 DfES 0034-2005 G Annesley, Nottingham: DfES Publications.

Dweck C (1999) *Self Theories: Their Role in Motivation, Personality and Development.* Hove, Sussex: Psychology Press.

Edwards A, Westgate D (1987) *Investigating Classroom Talk.* London: Falmer Press.

EPPI-Centre (2001) *EPPI-Centre Review Group Manual Version 1.1*. London: Social Science Research Unit, Institute of Education, University of London.

EPPI-Centre (2002) *EPPI-Centre Core Keywording Strategy: Data collection for a register of educational research. Version 0.9.5.* London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.

Feuerstein R, Rand Y, Hoffman MB, Miller R (1980) *Instrumental Enrichment: an Intervention Programme for Cognitive Modifiability.* Baltimore: University Park Press

Fullan M, Pomfret A (1977) Research on curriculum and instruction implementation. *Review of Educational Research* **47:** 335–397.

Gall MD, Borg WR, Gall JP (1996) *Educational Research: An Introduction* (6th ed.). London: Longman.

Galton G, Hargreaves L, Comber C, Pell A (1999) *Inside the Primary Classroom 20 Years On*. London: Routledge.

Garnham A, Oakhill J (1994) *Thinking and Reasoning.* Oxford: Blackwell.

Glass GV, McGaw B, Smith ML (1981) *Meta-Analvsis in Social Research*. Beverly Hills, California: Sage.

Greenhouse JB, Iyengar S (1994) Sensitivity analysis and diagnostics. In Cooper H, Hedges LV (eds) *The Handbook of Research Synthesis.* New York: Russell Sage, pages 383–398.

Hamers JHM, Van Luit JEH (1999) *Teaching and Learning Thinking Skills.* Abingdon: Swets and Zeitlinger.

Hamers JHM, Overtoom MT (eds) (1997) *Teaching thinking in Europe: Inventory of European Programmes*. Utrecht: SARDES.

Hattie JA (1992) Towards a model of schooling: a synthesis of meta-analyses. *Australian Journal of Education* **36**: 5–13.

Hattie J (1999) Influences on student learning. Inaugural lecture. Professor of Education, University of Auckland. August 2. Available at: http://www.arts.auckland.ac.nz/edu/staff/jhattie/Inaugural.html.

Hattie J (2004) Factors that influence children's learning: results from a study involving 500 meta-analyses. Paper presented at the ESRC Seminar 'Effective Educational Interventions', University of Newcastle upon Tyne, 8 July 2004.

Hattie J, Biggs J, Purdie N (1996) Effects of learning skills interventions on student learning: a meta-analysis. *Review of Educational Research* **66:** 99–136.

Hedges LV, Olkin I (1985) *Statistical Methods for Meta-Analysis*. San Diego, California: Academic Press.

Higgins S (2001) *Thinking Through Primary Teaching.* Cambridge: Chris Kington Publishing.

Higgins S, Baumfield V, Lin M, Moseley D, Butterworth M, Downey G, Gregson M, Oberski I, Rockett M, Thacker D (2004) Thinking skills approaches to effective teaching and learning: what is the evidence for impact on learners? In: *Research Evidence in Education Library.* London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.

Higgins S, Leat D (1997) Horses for courses or courses for horses: what is effective teacher development? *British Journal of In-Service Education* **23**: 303–314.

Higgins S, Moseley D (2002) Raising achievement in literacy through ICT. In: Monteith M (ed.) *Teaching Primary Literacy with ICT.* Buckingham: Open University Press, pages 30–45.

Kulik JA, Kulik C-LC (1989a) The concept of meta-analysis. *International Journal of Educational Research* **13:** 227–340.

Kulik JA, Kulik C-LC (1989b) Kulik JA, Kulik C-LC (1989) *Meta-analysis in Education*. Oxford, New York: Pergamon Press.

Leat D, Higgins S (2002) The role of powerful pedagogical strategies in curriculum development. *The Curriculum Journal* **13:** 71–85.

Lipman M (1991) *Thinking in Education* (1st ed.). Cambridge: Cambridge University Press.

Lipman M (2003) *Thinking in Education* (2nd ed.). Cambridge: Cambridge University Press.

Lipman M, Sharp A, Oscanyan F (1980) *Philosophy in the Classroom.* Princeton: Temple University Press.

Lipsey MW, Wilson DB (1993) The efficacy of psychological, educational, and behavioral treatment. *American Psychologist* **48**: 1181–1209.

Lipsey MW, Wilson DB (2001) *Practical Meta-analysis.* London: SAGE Publications.

MacLehose RR, Reeves BC, Harvey IM, Sheldon TA, Russell IT, Black AMS (2000) A systematic review of comparison of effect sizes derived from randomised and non-randomised studies. *Health Technology Assessment* **4 (34):** 1–154.

Marzano RJ (1998) *A Theory-Based Meta-Analysis of Research on Instruction*. Aurora, Colorado: Mid-continent Regional Educational Laboratory.

Marzano RJ, Pickering DJ, Pollock JE (2001) *Classroom Instruction That Works*. Alexandria Va: Association for Supervision and Curriculum Development (ASCD).

McGuinness C (1999) *From Thinking Skills to Thinking Classrooms: A Review and Evaluation of Approaches for Developing Pupils' Thinking*. Nottingham: DfEE Publications.

McGuinness C, Wylie J, Greer B, Sheehy NAF (1995) Developing children's thinking: a tale of three projects. *Irish Journal of Psychology* **16:** 378–388.

Mercer N (1995) *The Guided Construction of Knowledge: Talk Amongst Teachers and learners.* Clevedon: Multilingual Matters.

Moseley D, Baumfield V, Higgins S, Lin M , Miller J, Newton D, Robson S, Elliott J, Gregson M (2004) *Thinking Skill Frameworks for Post-16 Learners: An Evaluation.* A research report for the learning and skills research centre. Trowbridge: Cromwell Press.

Moseley D, Baumfield V, Elliott J, Higgins S, Miller J, Newton DP (2005a) *Frameworks for Thinking: A Handbook for Teachers and Learning.* Cambridge: Cambridge University Press.

Moseley D, Elliott J, Gregson M, Higgins S (2005b) Thinking skills frameworks for use in education and training. *British Educational Research Journal* **31**: 81–101.

Nickerson R, Perkins D, Smith E (1985) *The Teaching of Thinking*. London: Lawrence Earlbaum Associates.

Nisbet J, Davies P (1990) The curriculum redefined: learning to think-thinking to learn. *Research Papers in Education* **5:** 49–92.

Palincsar A, Brown A (1984) Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction* **1:** 117–175.

Romney DM, Samuels MT (2001) A meta-analytic evaluation of Feuerstein's Instrumental Enrichment program. *Educational and Child Psychology* **18:** 19–34.

Rosenshine B, Meister C (1994) Reciprocal teaching: a review of the research. *Review of Educational Research* **64:** 479–530.

Shayer M, Adey P (2002) *Learning Intelligence.* Buckingham: Open University Press.

Smith ML, Glass GV, Miller TI (1980) *The Benefits of Psychotherapy*. Baltimore, Maryland: Johns Hopkins University Press.

Sternberg RG, Bhana K (1986) Synthesis of research on the effectiveness of intellectual skills programs: snake-oil remedies or miracle cures? *Educational Leadership* **44:** 60–67.

Sipe T, Curlette WL (1997) A meta-synthesis of factors related to educational achievement: a methodological approach to summarizing and synthesizing meta-analyses. *International Journal of Educational Research* **25:** 583–698.

Torrance H, Pryor J (1998) *Investigating Formative Assessment: Teaching, Learning and Assessment in the Classroom.* Buckingham: Open University Press.

Trickey S, Topping KJ (2004) ‘Philosophy for children’: a systematic review. *Research Papers in Education* **19**: 335–380.

Walberg HJ (1984) Improving the productivity of America’s schools. *Educational Leadership* **41**: 19–27.

Webb N, Farivar S (1994) Promoting helping behaviour in co-operative small groups in middle school mathematics. *American Educational Research Journal* **31:** 369–396.

White R, Gunstone R (1992) *Probing Understanding.* London: Falmer Press.

Wilson V (2000) *Education Forum on Teaching Thinking Skills Report.* Edinburgh: Scottish Executive. Available from: http://www.scotland.gov.uk/library3/education/ftts-00.asp.

Wong BYL, Worsham AW, Austin GR (1985) Self-questioning instructional research: a review of effects of teaching thinking skills on SAT scores. *Review of Educational Research* **55:** 227–268.

Wood D, Wood H (1996) Vygotsky, tutoring and learning. *Oxford Review of Education* **22:** 5–16.

Appendix 1.1: Consultation and Critical Advisory Panel

Sue Eagle	Primary School Headteacher, Norfolk
Jane Brooke	LEA Adviser, Cheshire
Dr Carol McGuinness	Professor of Educational Psychology, Queen’s University, Belfast
David Thacker	Secondary Headteacher (retired)
Dr Iddo Oberski	Lecturer in Education, Stirling University
Dr Carolyn Tan	Senior Lecturer in Early Years Education, NIE, Nanyang Technological University (NTU), Singapore
Dr William Wu	Co-Director of the Thinking Qualities Initiative, Centre for Educational Development, Hong Kong Baptist University

Appendix 2.1: Inclusion and exclusion criteria

Criteria 1–5 were used in the initial literature search (Higgins *et al*., 2004). Criterion 6 was used to identify studies suitable for meta-analysis.

Inclusion criteria	**Exclusion criteria**
We included studies which:	We excluded studies which:
1. Are set in a school or schools and are concerned with any section of the school population (including pupils with special needs).	1. Are not set in a school or schools.
2. Evaluate the impact of the implementation of thinking skills interventions on teaching and/or learning. Thinking skills interventions are defined as approaches or programmes which require learners to articulate and evaluate learning strategies and/or which identify specific thinking processes that are amenable to instruction in order to improve teaching and/or learning. These interventions may be taught as separate programmes or infused into curriculum teaching. Impact includes: for example, pupil and/or teacher motivation and engagement, and/or patterns of classroom interaction, and/or self-regulation, and/or metacognitive monitoring, and/or pupil attainment.	2. Do not evaluate the impact of the implementation of thinking skills interventions on teaching and/or learning. Do not evaluate programmes or approaches which require the learners to articulate and evaluate the learning strategies that they are using and/or which do not identify specific thinking processes that are amenable to instruction in order to improve teaching and/or learning. Describe pupils' thinking or learning without any evaluation of a thinking skills intervention, strategy or approach. Do not evaluate the impact of thinking skills programmes and/or approaches.
3. Are concerned with the phases of compulsory schooling (5–16).	3. Are about pre-school, further and higher education, sixth form (A-level or equivalent).
4. Contain empirical classroom research with data or evidence (pupil outcomes, classroom processes, teacher role).	4. Are editorials, book reviews, policy documents, resources, guides, manuals, bibliographies, theoretical papers, philosophical papers, unevaluated interventions.
5. Are written in English[23].	5. Are not written in English.
6. Reported quantitative data on pupils' achievement or attitudes AND used a control or comparison group AND reported pre- and immediate post-intervention measures[24]; AND contained data on at least 10 pupils; AND reported an effect size for the intervention OR contained sufficient data to calculate an effect size of the impact of the thinking skills intervention or approach (e.g. means and SDs for pre- and post-test scores for intervention and comparison group); means, t-test values and group sizes; one-way ANOVA for the groups (with *F*-values and sample sizes); One-way ANCOVA with error values and correlation (*r*) between the covariate and dependent measure.	6. Do not contain data on pupil attainment or attitudes suitable for meta-analysis.

[23] It was beyond the funding of the review to translate papers from other languages. Although this may have restricted literature identified, every effort was made to identify studies from non-English speaking countries, but published in English.
[24] There were too few studies with suitable data which investigated retention or transfer effects, so for comparability, we only included studies with post-test data collected immediately after the intervention.

Appendix 2.2: Search strategy for electronic databases

Via BIDS

British Education Index (from 1986)
ERIC (Educational Resources Information Center) (from 1985)
IBSS (International Bibliography of the Social Sciences) (from 1980)
Ingenta Journals (full text of a large number of journals)
PsycINFO (extensive catalogue of psychology related publications)

Via Web of Science

Social Sciences Citation Index (SSCI) (from 1981)

Via FirstSearch

Article1st	Articles and tables of contents of journals in all subjects
Dissertations	Dissertation Abstracts, theses in all subjects
ECO	(Electronic Collections Online)
EducationAbs	Education Abstracts
PapersFirst	Conference papers in all subjects
Proceedings	Conference proceedings in all subjects
SIRS Researcher	Social Sciences
SocialSciAbs	Social sciences
WorldCat	Books and other materials on all subjects
Educati*on-line*	Conference papers and studies

Key search terms applied to each database were as follows:

thinking, thinking skills, thinking skills program(me), thinking strategies
critical thinking, critical thinking skills
creative thinking skills
higher order thinking skills (HOTS)
metacognition, metacognitive, meta-cognitive/ition
community of inquiry/enquiry/learners
transfer, near-transfer, far-transfer, bridging, teaching for transfer
reasoning, argument
Socratic questioning
mediated learning

The names of the following specific thinking skills programmes and approaches and their authors were also applied:

Instrumental Enrichment / Feuerstein
Somerset Thinking Skills / Blagg
Top Ten Thinking Tactics / Lake
Cognitive Acceleration in Science/Mathematics/Technology Education (CASE/CAME/CATE) / Adey, Shayer, Adhami
Philosophy for/with Children (P4C) / Lipman
Thinking Actively in a Social Context (TASC) / Wallace

Activating Children's Thinking Skills (ACTS) / McGuinness
CoRT (Cognitive Research Trust), Six Thinking Hats / de Bono
Storywise, Philosophy with Picture Books / Murris
Reason!Able / van Gelder

Appendix 2.3: EPPI-Centre keyword sheet, including review-specific keywords

V0.9.7 *Bibliographic details and/or unique identifier*

A1. Identification of report Citation Contact Handsearch Unknown Electronic database (Please specify.) .... **A2. Status** Published In press Unpublished **A3. 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.) .... **A4. Language** (Please specify.) .... **A5. In which country/countries was the study carried out?** (Please specify.)	**A6. 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.)......................... **A7. Curriculum** Art Business studies Citizenship Cross-curricular Design and technology Environment General Geography Hidden History ICT Literacy – first language Literacy further languages Literature Mathematics Music PSE Physical education Religious education Science Vocational Other (Please specify.).........................	**A8. Programme name** (Please specify.) .... **A9. 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.)............................ **A10. Age of learners** (years) 0–4 5–10 11–16 17–20 21 and over **A11. Sex of learners** Female only Male only Mixed sex	**A12. 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.).. **A13. 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

REVIEW SPECIFIC KEYWORDS

11. Pupil ages *How old were the pupils?* 5–*6* *11–12* *6–7* *12–13* *7–8* *13–14* *8–9* *14–15* *9–10* *15–16* *10–11* ***12. Teaching grouping*** How were the pupils grouped for teaching? *Usual class* *Set/Banded* *Mixed attainment/ability* *Not specified* *Special group* ***13. Teaching group size*** (Note: This might not be the same as Q15.) *Less than 15* *16–25* *26+* *Not recorded* **14. Teacher** *Who was the teacher?* Usual teacher Specialist/Expert Researcher as teacher (HEI staff) Teacher as researcher (school staff) (*Please specify.*) ______________________ **15. The research sample** How many schools were involved? _____ How many classes? ______. How many teachers involved? _____. How many pupils? ___. *(Intervention/Control)*	**16. Does the study sample focus on a particular group of learners?** All Special group Gifted and talented EAL Low attainers Other *(Please specify.)* _______________________ **17. Thinking skills terms** *Mark up to 3 categories for the main focus.* Argumentation Community of enquiry/learners Co-operative learning Creative thinking Critical thinking Decision making Discussion Enquiry based learning Higher order thinking Logical thinking Mediation/mediated learning Metacognition Problem solving Reflection Scaffolding Self-regulation Socratic questioning Systems thinking Transfer Others (*Please specify.*)_______________________ **18. Thinking skills approach** Infused Enrichment	**19. Type of data (***Mark all that apply.* Quantitative Qualitative Interactions Non-verbal behaviours Classroom talk/discourse Pupil attainment Pupil attitude/beliefs/dispositions Teacher attitude/beliefs/dispositions Other *(Pease specify.)*____________________ **20. Length of intervention** *(teaching time)* _______ lessons/hours *(Dlete as applicable.* Not recorded **21. Duration of intervention** *(from first lesson to last)* ______ weeks/months *(Dlete as applicable.*) Not recorded **22. Method of data collection** *Mark all that apply.* Observation Video Audio recording Test *(standardised, criterion referenced, SAT, GCSE,etc.*) Questionnaire/Survey/Rating scale Interview Document analysis Other *(Please specify.)*____________________	**Notes** v 0.1.1

Appendix 2.4: Calculating Hedges' *g*

The uways to calculate effect sizes are Glass's delta (the difference in the means of the two groups divided by the standard deviation of the *control group*), Cohen's d (the difference in the means of the two groups divided by the standard deviation of the *population*), and Hedges '*g*' (the difference in the means of the two groups divided by the standard deviation of the *sample*). (For a comprehensive discussion of the differences among various ways of estimating effect sizes, see Hedges and Olkin, 1985.)

Hedges' *g* is an inferential measure. It is normally computed by using the square root of the mean square error from the analysis of variance testing for differences between the two groups. It can also be calculated from other measures of effect size.

First, *d* (standardised mean difference) is calculated from the means, standard deviations and sample sizes:

$$d = \frac{\overline{X}_1 - \overline{X}_2}{s}$$

where:

$\overline{X}_1 = \textit{mean of intervention group}$

$\overline{X}_2 = \textit{mean of comparison or control group}$

$$s = \sqrt{\frac{(n_1 - 1)SD^2_1 + (n_2 - 1)SD^2_2}{n_1 + n_2 - 2}}$$

where:

$n_1 = \textit{sample size for group 1}$

$n_2 = \textit{sample size for group 2}$

$SD_1 = \textit{standard deviation group 1}$

$SD_2 = \textit{standard deviation group 2}$

d is then corrected for sample size using the formula:

$$g_c = g(1 - \frac{3}{4N - 9})$$

and its standard error is calculated:

$$SE = \sqrt{\frac{n_1 + n_2}{n_1 n_2} + \frac{(g_c)^2}{2(n_1 + n_2)}}$$

where: $n_1 = \textit{sample size of group 1 and}$

$n_2 = \textit{sample size of group 2}$

The heterogeneity statistic, *Q*, is calculated using the formula below. Since it is distributed as a chi-square, a p-value is obtained with *k* –1 degrees of freedom, where *k* is the number of effect sizes being combined.

$$Q = \sum w_i (\Theta_i - \Theta_{iv})^2$$

$$SE(\Theta_{IV}) = \frac{1}{\sqrt{\sum w_i}}$$

95% confidence intervals for individual and overall effects are calculated using the formula:

$$\Theta \pm (1.96 * SE(\Theta))$$

The test statistic (z) for overall effect is as follows:

$$z = \frac{\Theta}{SE(\Theta)}$$

Appendix 4.1: Summary of effects

Item ID	Item	Outcome	Group 1 size | mean | SD	Group 2 size | mean | SD	Hedges' g (corrected)	CI lower | upper
IT14198	Adey and Shayer (1990)[25]	Piagetian reasoning task	29 | 6.350 | 1.000	19 | 6.260 | 0.870	0.093	-0.486 | 0.672
IT16621	Adey *et al.* (2002)	Conservation test	122 | 4.050 | 2.950	66 | 2.530 | 2.250	0.555	0.251 | 0.860
IT16621	Adey *et al.* (2002)	Drawing test	302 | 13.930 | 4.580	166 | 12.170 | 5.500	0.357	0.166 | 0.548
IT13471	Cardelle-Elawar (1992)	Ravens 1	60 | 38.400 | 2.700	30 | 34.400 | 3.000	1.415	0.929 | 1.902
IT13471	Cardelle-Elawar (1992)	Ravens 2	30 | 39.400 | 3.500	32 | 35.800 | 3.000	1.093	0.557 | 1.630
IT13471	Cardelle-Elawar (1992)	ITBS 1	60 | 6.900 | 2.100	30 | 6.450 | 2.500	0.199	-0.240 | 0.638
IT13471	Cardelle-Elawar (1992)	ITBS 2	30 | 7.000 | 2.500	32 | 5.000 | 2.800	0.743	0.227 | 1.259
IT13471	Cardelle-Elawar (1992)	Mathematics achievement	60 | 15.300 | 1.670	30 | 6.150 | 1.190	5.935	4.946 | 6.924
IT13471	Cardelle-Elawar (1992)	Mathematics achievement 2	30 | 15.400 | 1.340	32 | 7.000 | 1.190	6.558	5.266 | 7.851
IT13471	Cardelle-Elawar (1992)	Attitude to mathematics 1	60 | 44.000 | 1.150	30 | 30.800 | 2.100	8.555	7.204 | 9.907
IT13471	Cardelle-Elawar (1992)	Attitude to mathematics 2	30 | 42.100 | 2.360	32 | 31.000 | 1.700	5.358	4.263 | 6.452
IT11895	Chang and Barufaldi (1999)	AFS concept	86 | 0.000 | 0.000	86 | 0.000 | 0.000	0.752	0.442 | 1.061
IT11895	Chang and Barufaldi (1999)	Science achievement	86 | 0.000 | 0.000	86 | 0.000 | 0.000	0.306	0.005 | 0.606
IT16404	Collings (1994)	GEFT1	20 | 0.000 | 0.000	41 | 0.000 | 0.000	1.076	0.341 | 1.811
IT16404	Collings (1994)	Science Reasoning 1	20 | 0.000 | 0.000	41 | 0.000 | 0.000	0.355	-0.183 | 0.894
IT16404	Collings (1994)	GEFT2	23 | 0.000 | 0.000	41 | 0.000 | 0.000	1.511	0.934 | 2.089
IT16404	Collings (1994)	Science Reasoning 2	23 | 0.000 | 0.000	41 | 0.000 | 0.000	0.909	0.373 | 1.445
IT16404	Collings (1994)	GEFT3	19 | 0.000 | 0.000	41 | 0.000 | 0.000	1.115	0.534 | 1.697
IT16404	Collings (1994)	Science Reasoning 3	19 | 0.000 | 0.000	41 | 0.000 | 0.000	0.671	0.113 | 1.229
IT13508	Csapó (1992)	Operational abilities	450 | 0.000 | 0.000	450 | 0.000 | 0.000	0.699	0.565 | 0.834
IT18183	Cunningham *et al.* (2002)	Attributional style	163 | -35.300 | 11.130	132 | -37.790 | 10.660	0.227	-0.003 | 0.458

[25] It should be noted that the impact of CASE appears to increase over time (Adey and Shayer 1993): this analysis uses the immediate post-test scores for comparability with other studies.

Item ID	Item	Outcome	Group 1 size | mean | SD	Group 2 size | mean | SD	Hedges' g (corrected)	CI lower | upper
IT18183	Cunningham *et al.* (2002)	coping efficacy	163 | -63.320 | 14.810	132 | -59.970 | 16.360	-0.215	-0.445 | 0.015
IT18183	Cunningham *et al.* (2002)	non-productive coping	163 | -43.030 | 11.430	132 | -47.810 | 13.740	0.381	0.149 | 0.613
IT18183	Cunningham *et al.* (2002)	non-productive coping – worry	163 | -29.650 | 13.970	132 | -34.610 | 15.710	0.335	0.104 | 0.566
IT18183	Cunningham *et al.* (2002)	non-productive coping – wishful thinking	163 | -55.940 | 19.030	132 | -60.530 | 21.650	0.226	-0.004 | 0.456
IT18183	Cunningham *et al.* (2002)	non-productive coping – not coping	163 | -40.040 | 15.560	132 | -44.410 | 19.910	0.247	0.017 | 0.478
IT18183	Cunningham *et al.* (2002)	non-productive coping – tension reduction	163 | -37.270 | 18.500	132 | -39.300 | 18.970	0.108	-0.121 | 0.338
IT18183	Cunningham *et al.* (2002)	non-productive coping – ignoring the problem	163 | -41.730 | 22.160	132 | -48.480 | 20.410	0.315	0.084 | 0.546
IT18183	Cunningham *et al.* (2002)	non-productive coping – self blame	163 | -38.650 | 25.570	132 | -43.210 | 25.880	0.177	-0.053 | 0.407
IT18183	Cunningham *et al.* (2002)	non-productive coping – keep to self	163 | -44.840 | 20.630	132 | -44.700 | 20.130	-0.007	-0.236 | 0.223
IT16665	De Koning and Hamers (1999)	Ravens	10 | 34.600 | 4.500	14 | 29.200 | 7.200	0.836	-0.016 | 1.687
IT16665	De Koning and Hamers (1999)	reading comprehension	10 | 112.000 | 12.000	14 | 99.900 | 7.400	1.223	0.327 | 2.118
IT16405	Donegan and Rust (1998)	Mc Daniel-Piers self-concept total	21 | 0.000 | 0.000	20 | 0.000 | 0.000	0.624	-0.005 | 1.253
IT15899	Greenberg (2000)	language test (school B)	63 | 0.000 | 0.000	46 | 0.000 | 0.000	1.004	0.601 | 1.408
IT15899	Greenberg (2000)	reading test (school B)	63 | 0.000 | 0.000	46 | 0.000 | 0.000	0.552	0.164 | 0.939
IT15899	Greenberg (2000)	mathematics test (school B)	63 | 0.000 | 0.000	46 | 0.000 | 0.000	1.002	0.599 | 1.406
IT15899	Greenberg (2000)	reading test (school C)	22 | 0.000 | 0.000	26 | 0.000 | 0.000	0.203	-0.366 | 0.772
IT15899	Greenberg (2000)	language test (school C)	22 | 0.000 | 0.000	26 | 0.000 | 0.000	0.899	0.301 | 1.497
IT15899	Greenberg (2000)	mathematics test (school C)	22 | 0.000 | 0.000	26 | 0.000 | 0.000	1.497	0.849 | 2.145
IT15899	Greenberg (2000)	reading test (school D)	58 | 0.000 | 0.000	44 | 0.000 | 0.000	0.307	-0.088 | 0.701
IT15899	Greenberg (2000)	language test (school D)	58 | 0.000 | 0.000	44 | 0.000 | 0.000	0.902	0.490 | 1.313
IT15899	Greenberg (2000)	mathematics test (school D)	58 | 0.000 | 0.000	44 | 0.000 | 0.000	1.099	0.678 | 1.520
IT13879	Haywood *et al*. (1988)	Ravens	26 | 0.000 | 0.000	27 | 0.000 | 0.000	0.673	0.118 | 1.228

Item ID	**Item**	**Outcome**	**Group 1 size | mean | SD**	**Group 2 size | mean | SD**	**Hedges' g (corrected)**	**CI lower | upper**
IT13879	Haywood *et al*. (1988)	PMA reasoning	26 | 0.000 | 0.000	27 | 0.000 | 0.000	1.201	0.612 | 1.790
IT13879	Haywood *et al*. (1988)	PMA spatial relations	26 | 0.000 | 0.000	27 | 0.000 | 0.000	0.747	0.189 | 1.306
IT13879	Haywood *et al*. (1988)	key math comprehension	26 | 0.000 | 0.000	27 | 0.000 | 0.000	-0.448	-0.993 | 0.098
IT13879	Haywood *et al*. (1988)	key math reasoning	26 | 0.000 | 0.000	27 | 0.000 | 0.000	-0.517	-1.066 | 0.031
IT13879	Haywood *et al*. (1988)	key math missing elements	26 | 0.000 | 0.000	27 | 0.000 | 0.000	-0.324	-0.866 | 0.219
IT13879	Haywood *et al*. (1988)	Peabody mathematics	26 | 0.000 | 0.000	27 | 0.000 | 0.000	-0.177	-0.716 | 0.363
IT13879	Haywood *et al*. (1988)	Metropolitan reading	26 | 0.000 | 0.000	27 | 0.000 | 0.000	-0.173	-0.712 | 0.367
IT13879	Haywood *et al*. (1988)	Stanford math application	26 | 0.000 | 0.000	27 | 0.000 | 0.000	-0.599	-1.151 | -0.048
IT16406	Hoek *et al.* (1999)	Mathematical reasoning	222 | 53.500 | 8.140	222 | 52.000 | 7.020	0.197	0.011 | 0.384
IT16406	Hoek *et al.* (1999)	Measures	222 | 34.300 | 10.500	222 | 30.300 | 10.300	0.384	0.196 | 0.572
IT16406	Hoek *et al.* (1999)	Information gathering	222 | 26.620 | 6.020	222 | 26.640 | 6.830	-0.003	-0.189 | 0.183
IT16647	Iqbal and Shayer (2000)	Piagetian test	124 | 7.100 | 0.000	94 | 0.000 | 0.000	1.066	0.780 | 1.353
IT16476	Kaniel and Reichenberg (1992)	Ravens	80 | 85.300 | 6.700	60 | 80.800 | 6.600	0.672	0.328 | 1.016
IT16476	Kaniel and Reichenberg (1992)	Figural analogies	80 | 93.900 | 5.060	60 | 69.300 | 23.700	1.533	1.152 | 1.914
IT16476	Kaniel and Reichenberg (1992)	Verbal analogies	80 | 88.800 | 11.100	60 | 78.300 | 14.300	0.831	0.482 | 1.180
IT16476	Kaniel and Reichenberg (1992)	Organiser	80 | 68.900 | 16.900	60 | 62.500 | 13.700	0.408	0.069 | 0.746
IT13839	Kramaski and Mevarech (1997)	graph construction	34 | 7.100 | 6.000	34 | 4.800 | 5.700	0.389	-0.092 | 0.869
IT13839	Kramaski and Mevarech (1997)	information processing	34 | 5.110 | 0.600	34 | 3.250 | 0.520	3.275	2.535 | 4.015
IT13839	Kramaski and Mevarech (1997)	error detection	34 | 1.300 | 0.260	34 | 0.580 | 0.230	2.900	2.208 | 3.591
IT13839	Kramaski and Mevarech (1997)	social cognitive interaction	34 | 1.900 | 0.210	34 | 1.000 | 0.190	4.443	3.539 | 5.348
IT13839	Kramaski and Mevarech (1997)	extrapolation	34 | 1.200 | 0.240	34 | 1.090 | 0.220	0.472	-0.010 | 0.955
IT13839	Kramaski and Mevarech (1997)	logo error detection	34 | 1.600 | 0.840	34 | 1.420 | 1.020	0.190	-0.286 | 0.667
IT13839	Kramaski and Mevarech (1997)	Self-reflection	34 | 13.000 | 3.430	34 | 9.830 | 2.120	1.099	0.587 | 1.611
IT16087	Maqsud (1998)	Ravens	20 | 36.600 | 2.930	20 | 34.000 | 3.210	0.829	0.181 | 1.478
IT16087	Maqsud (1998)	metacognitive awareness	20 | 39.800 | 5.010	20 | 34.800 | 5.130	0.967	0.308 | 1.625
IT16087	Maqsud (1998)	mathematics 1	20 | 11.700 | 2.370	20 | 8.640 | 2.620	1.201	0.522 | 1.880
IT16087	Maqsud (1998)	mathematics 2	20 | 10.700 | 1.980	20 | 6.850 | 2.240	1.785	1.041 | 2.529
IT16087	Maqsud (1998)	mathematics 3	20 | 12.100 | 2.310	20 | 9.370 | 1.980	1.244	0.561 | 1.927

Item ID	**Item**	**Outcome**	**Group 1 size | mean | SD**	**Group 2 size | mean | SD**	**Hedges' g (corrected)**	**CI lower | upper**
IT16087	Maqsud (1998)	mathematics 4	20 | 11.800 | 2.860	20 | 8.360 | 2.640	1.225	0.544 | 1.906
IT16087	Maqsud (1998)	mathematics achievement	20 | 38.300 | 7.540	20 | 31.900 | 6.870	0.870	0.218 | 1.521
IT16087	Maqsud (1998)	attitude to mathematics	20 | 72.500 | 10.400	20 | 64.400 | 11.300	0.731	0.089 | 1.373
IT16229	Martin (1984)	Ravens	9 | 0.000 | 0.000	9 | 0.000 | 0.000	0.872	-0.107 | 1.850
IT16229	Martin (1984)	SAT-HI mathematics	9 | 0.000 | 0.000	9 | 0.000 | 0.000	0.524	-0.420 | 1.468
IT13613	Mercer *et al*. (1999)	Ravens	60 | 0.000 | 0.000	64 | 0.000 | 0.000	0.314	-0.040 | 0.668
IT16478	Muttart (1984)	primary mental abilities	9 | 166.890 | 15.740	8 | 153.630 | 11.870	0.895	-0.118 | 1.907
IT16478	Muttart (1984)	basic skills	9 | 267.000 | 43.650	8 | 231.750 | 18.740	0.974	-0.049 | 1.997
IT16478	Muttart (1984)	self-concept of ability	9 | 26.330 | 4.000	8 | 26.250 | 5.470	0.016	-0.936 | 0.968
IT16478	Muttart (1984)	academic self-concept	9 | 6.330 | 1.500	8 | 7.380 | 2.450	-0.498	-1.470 | 0.473
IT16478	Muttart (1984)	achievement self-esteem	9 | 20.000 | 2.400	8 | 19.130 | 1.460	0.409	-0.556 | 1.375
IT16478	Muttart (1984)	Self-concept	9 | 83.890 | 2.930	8 | 78.500 | 5.560	1.174	0.120 | 2.227
IT16054	Oladunni (1988)	creative mathematics problems	76 | 0.000 | 0.000	85 | 0.000 | 0.000	1.029	0.700 | 1.359
IT16408	Riding and Powell (1987)	Ravens high achieving boys	8 | 2.500 | 1.800	8 | 2.630 | 2.740	-0.053	-1.033 | 0.927
IT16408	Riding and Powell (1987)	Ravens low achieving boys	8 | 6.130 | 3.300	8 | 2.000 | 2.060	1.419	0.288 | 2.551
IT16408	Riding and Powell (1987)	Ravens high achieving girls	8 | 2.880 | 1.760	8 | 1.000 | 2.650	0.790	-0.239 | 1.820
IT16408	Riding and Powell (1987)	Ravens low achieving girls	8 | 3.380 | 2.290	8 | 1.500 | 2.920	0.677	-0.339 | 1.694
IT16408	Riding and Powell (1987)	Reading high achieving boys	8 | 30.500 | 11.800	8 | 29.800 | 12.000	0.056	-0.925 | 1.036
IT16408	Riding and Powell (1987)	Reading high achieving girls	8 | 27.000 | 13.800	8 | 29.300 | 11.800	-0.169	-1.152 | 0.813
IT16408	Riding and Powell (1987)	Reading low achieving boys	8 | 23.600 | 4.820	8 | 13.800 | 8.740	1.313	0.202 | 2.424
IT16408	Riding and Powell (1987)	Reading low achieving girls	8 | 27.500 | 12.300	8 | 14.600 | 9.660	1.103	0.029 | 2.177
IT16408	Riding and Powell (1987)	Mathematics high achieving boys	5 | 13.400 | 8.160	5 | 18.000 | 8.880	-0.487	-1.757 | 0.782
IT16408	Riding and Powell (1987)	Mathematics low achieving boys	5 | 16.400 | 3.070	5 | 13.200 | 5.150	0.682	-0.616 | 1.979
IT16408	Riding and Powell (1987)	Mathematics high achieving girls	5 | 22.200 | 6.520	5 | 27.600 | 6.120	-0.771	-2.085 | 0.542
IT16408	Riding and Powell (1987)	Mathematics low achieving girls	5 | 16.000 | 4.150	5 | 20.800 | 6.240	-0.818	-2.141 | 0.504

Item ID	Item	Outcome	Group 1 size | mean | SD	Group 2 size | mean | SD	Hedges' g (corrected)	CI lower | upper
IT13887	Ritchie and Edwards (1996)	Torrance fluency	22 | 44.900 | 24.100	18 | 40.800 | 16.700	0.190	-0.434 | 0.815
IT13887	Ritchie and Edwards (1996)	Torrance flexibility	22 | 24.400 | 9.330	18 | 21.700 | 6.100	0.329	-0.299 | 0.956
IT13887	Ritchie and Edwards (1996)	Torrance originality	22 | 24.250 | 16.500	18 | 24.200 | 11.600	0.003	-0.620 | 0.626
IT16660	Schmid and Telaro (1990)	Reading 1	8 | 50.000 | 14.800	4 | 44.500 | 5.740	0.397	-0.818 | 1.613
IT16660	Schmid and Telaro (1990)	Reading 2	10 | 60.700 | 16.300	7 | 54.100 | 11.200	0.433	-0.547 | 1.413
IT16660	Schmid and Telaro (1990)	Reading 3	6 | 75.500 | 15.700	8 | 71.800 | 14.800	0.228	-0.835 | 1.291
IT16407	Naval-Severino (1993)	Torrance fluency	7 | 0.000 | 0.000	8 | 0.000 | 0.000	1.409	0.237 | 2.581
IT16407	Naval-Severino (1993)	Torrance flexibility	7 | 0.000 | 0.000	8 | 0.000 | 0.000	1.381	0.215 | 2.548
IT16407	Naval-Severino (1993)	Torance originality	7 | 0.000 | 0.000	8 | 0.000 | 0.000	1.209	0.076 | 2.341
IT16484	Shayer and Beasley (1987)	PMA verbal	10 | 0.900 | 0.700	10 | 1.200 | 0.800	-0.382	-1.269 | 0.504
IT16484	Shayer and Beasley (1987)	PMA spatial	10 | 0.900 | 2.400	10 | 0.300 | 2.150	0.252	-0.629 | 1.133
IT16484	Shayer and Beasley (1987)	Neale reading accuracy	10 | 0.500 | 0.650	10 | 0.300 | 0.550	0.318	-0.565 | 1.202
IT16484	Shayer and Beasley (1987)	Neale reading rate	10 | 1.200 | 0.650	10 | 0.900 | 0.750	0.409	-0.478 | 1.297
IT16484	Shayer and Beasley (1987)	Piagetian battery	10 | 1.700 | 1.400	10 | 0.000 | 1.200	1.249	0.272 | 2.226
IT13507	Strang and Shayer (1993)	Chemistry test	11 | 0.000 | 0.000	10 | 0.000 | 0.000	1.116	0.181 | 2.050
IT15898	Tenenbaum (1986)	Algebra test (CPR =FB/C)	30 | 88.270 | 9.860	30 | 63.170 | 14.700	1.979	1.355 | 2.604
IT15898	Tenenbaum (1986)	Algebra test (mastery learning)	28 | 77.890 | 12.430	30 | 63.170 | 14.700	1.064	0.511 | 1.616
IT15898	Tenenbaum (1986)	Science test (CPR+FB/C)	29 | 85.280 | 58.710	31 | 11.720 | 15.150	1.720	1.121 | 2.318
IT15898	Tenenbaum (1986)	Science test (mastery learning)	30 | 73.970 | 12.930	31 | 58.710 | 15.150	1.068	0.529 | 1.607
IT16670	Tzuriel and Alfassi (1994)	Ravens	93 | 0.000 | 0.000	98 | 0.000 | 0.000	0.131	-0.153 | 0.415
IT16670	Tzuriel and Alfassi (1994)	RSDT	93 | 0.000 | 0.000	98 | 0.000 | 0.000	0.231	-0.053 | 0.516
IT16670	Tzuriel and Alfassi (1994)	Organiser	93 | 0.000 | 0.000	98 | 0.000 | 0.000	0.256	-0.029 | 0.541
IT13466	Ward and Traweek (1993)	Reading – word identification	11 | 54.900 | 9.000	12 | 43.000 | 13.000	1.017	0.138 | 1.897
IT13466	Ward and Traweek (1993)	Reading – passage comprehension	11 | 79.100 | 15.500	12 | 49.300 | 19.700	1.611	0.647 | 2.576

Appendix 4.2: Details of studies included in the in-depth review

In-depth review studies: educational setting, intervention type, study type and results

Item ID	Item	Educational setting	Programme or intervention type	Study type	Results of the study as reported by authors
IT14198	Adey and Shayer (1990) Accelerating the development of formal thinking in middle and high school students	Secondary school 'The main experiment schools...Two were middle schools (9–14 years) and the remainder were secondary schools' (p 270#837)	Yes: *Cognitive Acceleration Through Science Education (CASE)*	Evaluation: Researcher-manipulated *Intervention study with pre-post tests and control group*	'Results of tests given immediately after the intervention and 1 and 2 years later, standardised with respect to pre-test scores of experimental and control pupils, indicate that the intervention led to immediate gains in Piagetian measures of cognitive development and to gains in experimental groups' achievement in science, mathematics and English language measured 2 and 3 years after intervention programme. Groups most affected included the boys who started the programme in Year 8 (Grade 7) and the girls who started in Year 7 (Grade 8). There were stronger effects on girls' gains in English achievement and on boys' gains in science and mathematics achievement.' (p 1 #747)
IT16621	Adey *et al.* (2002) Effects of a cognitive accleration programme on Year 1 pupils	Primary school 10 experimental primary schools (Year 1 classes) and five controls in the same inner London LEA (p 1)	Yes: *Cognitive Acceleration (CA)*	Evaluation: Researcher-manipulated *Intervention study with pre-post tests and control group*	'The simplest question to ask about the effect of the intervention is "Did it work, in the sense of increasing the rate of cognitive development in the CA classes relative to the control classes?! The answer is yes, it did."' (p 13) Effect sizes 0.43 and 0.47 (conservation and drawing). Main effect for teacher on post-scores.

IT13471	Cardelle-Elawar (1992) Effects of Teaching metacognitive Skills to Students with Low Mathematical Ability	Primary school US 6th grade in a public elementary school (p 112)	No: *Metacognitive skills*	Evaluation: Researcher-manipulated *Intervention study with pre-post tests and control group. Follow-up investigation with new control group to eliminate confounding variable the first phase when the researcher was also the instructor for the experimental group.*	Phase 1 (p 116) Table 1 and reporting of ANCOVA analysis shows that there were significant differences between the pre- and post-tests for the experimental group and that they scored higher than the control group on all the measures post-test. Phase 2 Table 1 and ANCOVA scores (p 117) and Table 2 (p 118) show that the experimental group had higher scores than the control group in all post-test measures and these are statistically significant. Discussion refers to analysis of improvements in aspects of problem solving and says that students have got better at understanding how to approach a problem, identifying the appropriate schema for organising the information, recognising that there may be more than one way to solve a problem and verifying their solutions. Improvements are interpreted as arising from increasing their linguistic comprehension of key words and sentences leading to heightened concentration and reduced impulsivity and an ability to reflect on their own thinking. (p 119)
IT11895	Chang and Barufaldi (1999) The use of a problem-solving-based instructional model in initiating change in students' achievement and alternative frameworks	Secondary school Public Junior High School	Yes: *Search, Solve, Create, Share (SSCS) (Pizzini et al., 1988)*	Evaluation: Researcher-manipulated *Four classes taught by the same teacher participated. Researcher randomly assigned students from two intact classes to the control group and from two intact classes to the experimental group. Cluster randomised controlled trial.*	Problem-based instructional method produced significantly greater achievement of 9th grade earth science students than did conventional teaching method especially at the application level (Table 3, p 381). It also statistically supported that students in experimental group experienced a significant conceptual change (Table 4 p 381) even though both the intervention and traditional methods were successful in modifying AFs. Student opinion in experimental group showed no particular perceptions toward the teaching approach but did express advantages in terms of helping them develop analytical and observing skills and improving their thinking skills – pupil quotes (pp 381–382). Pupils had concerns of value of the approach given pressure of exams.
IT16404	Collings (1994) Some fundamental questions about scientific thinking	Secondary school in a rural setting	No: *'a blend of field independence, scientific reasoning and metacognition' (p 169)*	Evaluation: Researcher-manipulated *Intervention study with pre-post tests and control group*	'significant increases over controls had taken place in all groups undergoing training' (p 172)

IT13508	Csapó (1992) Improving Operational Abilities in Children	Primary school Secondary school	Not stated/unclear : *Theoretical basis is clearly stated as neo-Piagetian, but no specific programme name is given.*	Evaluation: Researcher-manipulated	The structured task systems in the intervention has different effects across the various training contexts. *Systematising ability*: develops rapidly across the age range studied but is not accelerated by the intervention and tasks designed to improve it had only a weak effect on other abilities. *Logical operations*: develop slowly but can be improved at the younger age. Intervention also had a significant effect on the other abilities at an older age. The intervention did improve thinking but this change was not detected by the measurement used (possible ceiling effect) – suggests that improvement in advances formal thinking may not be characterised in terms of formal logic. *Combinative operations*: develop at an intermediate pace and the intervention achieved considerable acceleration at both ages. Differences between these three abilities are greater than their similarities. Effects less in groups where the intervention applied in more than one subject. Enrichment materials to improve logical and combinative ability are worthwhile but less certain for systematising ability which appears already well covered by normal teaching and is not significantly improved by the tasks. (p 157)
IT18183	Cunningham *et al*. (2002) Enhancing coping resources in early adolescence through a school-based program teaching optimistic thinking skills	Secondary school 16 year 5 and 6 classes	Yes: *Bright Ideas: Skills for positive thinking*	Evaluation: Researcher-manipulated *Intervention study with pre-post tests and control group*	'program children reported significant increases in coping efficacy, together with significant reductions in depressive attributions and utilization of non-productive coping strategies' (p 376). 'Compared to controls, children who participated in the program reported reduced use of the non-productive coping strategies of worry, wishful thinking, not coping and ignoring the problem' (p 377).
IT16665	De Koning and Hamers (1999) Teaching inductive reasoning: theoretical background and educational implications	Primary school	Yes: *Inductive reasoning*	Evaluation: Researcher-manipulated *Intervention study with pre-post tests and control group*	The authors report that 'the teacher was able to deal with the main didactical requirements of the programme' (p 180). Effect sizes reported: Ravens –0.84 and reading comprehension 1.01.
IT16405	Donegan and Rust (1998) Rational emotive education for improving self concept in second grade students	Primary school in rural Tennessee	Yes: *Vernon's Thinking, Feeling, Behaving*	Evaluation: Researcher-manipulated *Intervention study with pre-post tests and control group*	Total McDaniels scores revealed a significant difference between the experimental group's pre- and post-tests and between the experimental and control pre-tests. Analysis of subscales and BASE scores revealed no other significant relationships however.

IT15899	Greenberg (2000) Attending to hidden needs: the Cognitive Enrichment Advantage Perspective	Primary school	Yes: *COGNET/CEA: Cognitive Enrichment Advantage*	Evaluation: Researcher-manipulated *Project COGNET designed by evaluators*	'the results... revealed educationally significant effect size in favour of COGNET treatment groups over their matched comparison groups in all three subject areas' (p 66). ES range: Reading: 0.2-.55, Language: 0.9-1 Mathematics: 1-1.5
IT13879	Haywood *et al.* (1988) Cognitive education with deaf adolescents: effects of Instrumental Enrichment	Special needs school 'two public residential schools...for deaf students' (p 28)	Yes: *Feuerstein's Instrumental Enrichment*	Evaluation: Researcher-manipulated *Intervention study with pre-post tests and control group. The study was funded as a demonstration project, but the researchers were not in position to control all relevant variables.*	As anecdotal quotations in the text (pp 38–39), in textual form and in tables 2, 3, 4 and 5 (pp 31–34). Significant gains were made by the experimental group on 6 out of 7 of the mastery, transfer and IE vocabulary tests. The IE students gained significantly more than control students on all three aptitude tests (Raven's Matrices and PMA reasoning and spatial relations). The control students gained significantly more than the IE students on Stanford Mathematics applications, and (although the differences were not statistically significant) did better in absolute terms on 7 out of the 10 other measures of scholastic achievement. The multiple discriminant analysis correctly identified 75% of the high and low gain groups, using four steps: sex, group (IE versus control), number of IE instruments taught and race. It is not stated how each variable was related to gain. Informal observations by the teachers and investigators, and comments gathered by one teacher suggested that IE helped improve motivation to learn, attitude towards learning and attitude towards the self as a learner. Anecdotal reports from teachers revealed indicated improvements in social behaviour as well as in academic performance and generally enthusiastic responses.
IT16406	Hoek *et al.* (1999) The effects of integrated social and cognitive strategy instruction on the mathematics achievement in secondary education	Secondary school	Yes: *AGO (Dutch acronym = adaptive instruction and co-operative learning)*	Evaluation: Researcher-manipulated *Intervention study with pre-post tests and control group*	The authors report cautious support for H1 (programme effect), which is supported by the ANOVA but not the multilevel analysis. H2 (differential effect) is supported in relation to mathematic reasoning, inconclusive in respect of measurement and rejected in terms of information gathering. H3 (remedial effect) is only supported in terms of information gathering. The authors conclude that the complexity of these results mean that there can be positive effects for both high and low achieving students from this kind of intervention, although combining strategy and co-operative working benefited able students more due to the higher cognitive demands of the integrated programme.

IT16647	Iqbal and Shayer (2000) Accelerating the development of formal thinking in Pakistan secondary school students: achievement effects and professional development issues	Secondary school Private school A (female) Government school B (male) Private school C (male and female)	Yes: *Cognitive Acceleration in Science Education (CASE)*	Evaluation: Researcher-manipulated *Intervention study with pre-post tests and comparison data*	'where conditions were favourable, and where teachers were able and willing to develop new professional skills. Use of CASE methodology produced cognitive gains of the order of 1 standard deviation, and brought students up some 30 percentile points on representative population norms.' (p 271) Cognitive effects: overall 1.28; School A: 0.91, School B: 1.67, School C: 0.75. Science achievements: School A: 0.7, School B: 0.37, School C: 0.41.
IT16476	Kaniel and Reichenberg (1992) Instrumental Enrichment – effects of generalisation and durability with talented adolescents	Secondary school in deprived areas	Yes: *Instrumental Enrichment*	Evaluation: Researcher-manipulated	'The results show clearly that with most ages, immediately after the IE program was completed, there was an increase in performance on the task. However there was no effect on the non-verbal thinking task (suggested ceiling effect)' (p 131). For the four-year follow-up, 'the IE group demonstrated higher scores in verbal and non-verbal thinking and in achievement' (p 133).
IT13839	Kramarski and Mevarech (1997) Cognitive-metacognitive training within a problem-solving based LOGO environment	Secondary school Integrated junior high school in Israel	Yes: *Metacognitive Strategy use of LOGO Learning Environment*	Evaluation: Researcher-manipulated *Intervention study with pre-post tests and control group*	'Overall, results indicate that students in the co-operative setting significantly outperformed their counterparts in the individualised setting and in the non-treatment group on the two measures of originality and on verbal flexibility' (p 107). Individual Logo use produced a significant difference in originality between this group and the control. However, the authors emphasise the superiority of the co-operative approach, in interpersonal terms as well as creative ones.
IT16087	Maqsud (1998) Effects of metacognitive instruction on mathematics achievement and attitude towards mathematics of low mathematics achievers	Secondary school	Yes: *Metacognition*	Evaluation: Researcher-manipulated *Intervention study with pre-post tests and control group*	'The posttest scores of the experimental group for all four variables were found significantly higher than those for the control group' (p 242).

IT16229	Martin (1984) Cognitive modification for the hearing impaired adolescent: the promise	Special needs school Model Secondary School for the Deaf, Washington DC	Yes: *Instrumental Enrichment (IE)*	Evaluation: Researcher-manipulated *Intervention study with pre-post tests and control group*	Hypothesis 1: That the experimental group would show greater gains as measured by Raven's against the control was said to be supported, although the difference between the groups did not achieve significance (p<0.07). Hypothesis 2: That the experimental group would show greater gains as measured by reading comprehension SAT-HI against the control was not upheld, non-significant difference on t test, although the author claims significance for the gain for the experimental group. Problem-solving interview data showed a trend favouring the experimental group but there was no year 2 follow-up.
IT13613	Mercer *et al.* (1999) Children's talk and the development of reasoning in the classroom	Primary school 'state middle schools' (p 101)	Yes: *The TRAC (Talk, Reasoning and Computers) programme, subsequently published as 'Thinking Together'*	Evaluation: Researcher-manipulated *Intervention study with pre-post tests and control group*	'We therefore have evidence that the use of exploratory talk helps the joint solution of problems' (p 105). 'We have also shown that the intervention programme increased the amount of exploratory talk used by the target focal groups' (p107). 'It can be seen that there is a relatively greater improvement in the scores for the target classes, which is in accord with our hypothesis' although 'not statistically significant' p107 'The gains made by the individual target class children were significantly greater than those made by children in control classes' (p 107): (a) using the kind of language we call 'exploratory talk' helps children to work more effectively together on problem solving tasks. (b) using a specially designed programme of teacher-led and group-based activities, teachers can increase the amount of exploratory talk used by children working together in the classroom; and (c) children who have been taught to use more exploratory talk make greater gains in their individual scores on the Raven's test of reasoning than do children who have had no such teaching (p108). The intervention programme increased the amount of 'exploratory talk' used by the target focal groups when solving Raven's Matrices reasoning problems (Tables IIb and IIIb, pp 106–107). (Please note that this finding is as reported by the authors, but is based on an error in interpreting the supporting statistical analyses.) Adherence to the ground rules helped groups solve the reasoning test problems, as it was found that when arriving at correct group solutions, there was a high level of 'exploratory talk' in one group of three children (Table I, p 105). Target children's individual (but not group) performance on Raven's Matrices improved (Tables IV and V, p 107)

IT16478	Muttart (1984) Assessment of Effects of Instrumental Enrichment Cognitive Training	Secondary school Students in remedial programmes	Yes: *Feuerstein Instrumental Enrichment*	Evaluation: Researcher-manipulated *Intervention study with pre-post tests and control group*	Significant positive differences in achievement and self concept as well as verbal abilities
IT16407	Naval-Severino (1993) Developing creative thinking among intellectually able Filipino children from disadvantaged urban communities	Other early years setting Daycare settings	Yes: *Creative thinking*	Evaluation: Researcher-manipulated *Intervention study with pre-post tests and comparsion group*	'Results of t-test analyses revealed significant differences between the pre and post test scores of the children only for Group B which received more training sessions' (p 122).
IT16054	Oladunni (1998) An experimental study on the effectiveness of metacognitive and heuristic problem solving techniques on computational performance of students in mathematics	Secondary school	Yes: *Metacognitive and heuristic techniques*	Evaluation: Researcher-manipulated *Intervention study with pre-post tests and control group*	HO1: That there is no difference between performance of experimantal and control groups – rejected HO2: That there is no difference between performance of ability groups – rejected HO3: That there is no difference in performance by gender – upheld
IT16408	Riding and Powell (1987) The effect on reasoning, reading and number performance of computer-presented critical thinking activities in five year old children	Primary school In an urban area	Yes: *Critical thinking skills*	Evaluation: Researcher-manipulated	Performance on Raven's – significant effect of intervention Performance on Reading – significant effect of intervention and pre-test performance Performance on Mathematics – significant effect of gender (girls performed better)

IT13887	Ritchie and Edwards (1996) Creative thinking instruction for aboriginal children	Secondary school	Yes: *CoRT (Cognitive Research Trust) Thinking programme*	Evaluation: Researcher-manipulated *Intervention study with pre-post tests and control group. The researcher trained three teachers in three different schools to deliver CoRT lessons. Two of the three control classes were in other schools, as there were not enough classes in the same schools with a high proportion of Aboriginal students.*	The CoRT lessons did not significantly affect cognitive ability, teacher-rated school achievement, self-reported use of CoRT thinking approaches, self-concept as a thinker or internal locus of control. Significant and substantial overall gains were made on the Torrance Test measures (fluency, flexibility and originality). Trend analysis showed a consistent upwards movement in fluency and flexibility, but a levelling off after an initial gain in originality. Implementation integrity was satisfactory, although there were some problems in achieving effective group work. The teachers' familiarity and acceptance of the CoRT approach was found wanting in two respects: they felt uneasy about using CoRT skill acronyms and they did not always demonstrate enthusiasm and confidence. However, students experienced a high level of success in the lessons.
IT16685	Scheinin and Mehtäläinen (1999) Applying the theory of knowledge to teaching thinking	Secondary school	Yes: *FACE project (Formal Aims of Cognitive Education) with an philosophical/ epistemological focus*	Evaluation: Researcher-manipulated *'Quasi-experimental combination of follow-up and cohort study' (p 91)*	'The intervention effects on performance of tests of cognitive ability were nonsignificant. It may be noted however that all of the tests of cognitive ability except for the Box Folding test, showed a positive but nonsignificant difference in favour of the experimental group' p 96 'Positive intervention effects were evident in the tests measuring formal cognitive skills. In both tests the experimental group improved its performance beyond the age-typical performance of the controls' p 97–98. 'In the post-test the self concept of the experimental group was clearly more positive than that of the the control goup on the dimension of orderliness of thinking, mathematical ability and linguistic ability.'
IT16660	Schmid and Telaro (1990) Concept mapping as an instructional strategy for high school biology	Secondary school	Yes: *Concept mapping*	Evaluation: Researcher-manipulated *Intervention study with pre-post tests and control group*	'The results of this study reflect positively on the implementation of concept mapping in high school biology, but not without qualifications' (p 82). Less capable students did better at higher levels of learning after concept mapping but for high achievers the concept mapping was 'disruptive' and did not produce large effects (ES 0.23, compared with ES 0.43 for the medium and ES 0.40 for the low achievers groups). Overall, affective responses to concept mapping were lukewarm, even from the groups who benefited most.
IT16484	Shayer and Beasley (1987) Does Instrumental Enrichment Work?	Secondary school	Yes: *Feuerstein's Instrumental Enrichment (IE)*	Evaluation: Researcher-manipulated	Effects were greatest on measures of 'fluid intelligence' (Piagetian and Ravens) with ES of 1.22 and 1.07. The results from 'crystallised intelligence' and achievement measures were more disappointing: 'These are not gains which commend themselves for emulation' (p 111).

IT13507	Strang and Shayer (1993) Enhancing high school students' achievement in chemistry through a thinking skills approach	Secondary school Setting is described as 'a London Comprehensive School' (p 323).	Yes: *Feuerstein's Instrumental Enrichment (IE)*	Evaluation: Researcher-manipulated *A class of students was divided into two groups with one group receiving lessons designed to compensate for previously identified cognitive deficits, and the other group receiving normal lessons taught by the head of science.*	The authors report a significant difference between the post-test scores of the experimental and control group. The results of the regression equation suggest a bimodal distribution with some participants benefiting more than others. This means that the t-test score is misleading as most of the effect is located in four high scoring participants. The result of the experiment is not uniform for the experimental group; the reason for the effect appears to lie neither in the gender of the student nor in their initial abilities.
IT15898	Tenenbaum (1986) The effect of quality of instruction on higher and lower mental processes and in the prediction of summative assessment	Secondary school	Yes: *Cues, participation and reinforcement with feedback corrective learning and mastery learning*	Evaluation: Researcher-manipulated	'The differences between the enhanced CPR+FB/C groups and the conventional group on both LMP and HMP were found to be statistically significant (p<0.5). The mean final achievement of the enhanced CPR+FB/C groups were about 15% above the conventional groups on LMP and 36% on HMP. The mastery learning groups also demonstrated significant (p<0.5) advantage over the conventional groups on lower and higher mental process.' (p 110) Enhanced CPR+FB/C groups ES 1.71 (9th grade) 1.75 (6th grade). Mastery learning ES 1 (9th grade) 1 (6th grade)
IT16670	Tzuriel and Alfassi (1994) Cognitive and motivational modifiability as a function of the instrumental enrichment (IE) program	Secondary school	Yes: *Feuerstein's Instrumental Enrichment (IE)*	Evaluation: Researcher-manipulated *Intervention study with pre-post tests and control group*	The scores on the SVII element of the LPAD improved more for the non-IE group, RDST element showed no differences and only the Organiser test showed advantage from the IE group.. 'there is a correspondence in the nonprogram group between cognitive modifiability before and after IE. Subjects with low, medium and high Initial Cognitive Modifiability (pre-IE) were respectively low, medium and high on a different measure of cognitive modifiability two years later. In the IE program group, on the other hand, the highest gains at the end of IE were among the subjects with low Initial Cognitive Modifiablity, and less among medium and high Initial Cognitive Modifiablity subjects.' (p 113)
IT13466	Ward and Traweek (1993) Application of a Metacognitive Strategy to Assessment, Intervention, and Consultation: a Think-Aloud Technique: Study 1	Primary school fifth-grade with mean age of 10 years 10 months, (p 473)	Yes: *A think-aloud procedure (Meyers and Lytle, 1986; Meyers and Wade 1990)*	Evaluation: Researcher-manipulated *Experimental and control groups with students assigned on the basis of reading test*	'Students who followed the think-aloud procedure requiring them to talk about their reading strategies as they performed a cloze task improved their reading comprehension scores significantly more than those who performed the cloze task without using the think-aloud procedure (word identification: mean scores of 54.9 against 43.0; mean scores passage comprehension: 79.4 against 49.3' (p 475). 'Five of the ten questions used during reading that were believed to be most salient in generating general metacognitive strategies were analysed and no differences were found between the groups' (p 476).