

Picking the Strawberries out of the Jam: thinking critically about Systematic Reviews and Meta-Analysis

Steve Higgins

Elaine Hall

Centre for Learning and Teaching

School of Education, Communication and Language Sciences,

University of Newcastle upon Tyne.

Abstract:

Educational research has as one of its goals the provision of research evidence to support or change practice and in recent years there has been considerable debate over what constitutes 'good evidence'. Broadly speaking, there has been an increasing focus on research methods which have apparent claims to greater objectivity – such as systematic reviewing and meta-analysis – when compared to the narrative reviews previously more common in the educational research paradigm. This paper draws on our experience of conducting systematic reviews of the impact of thinking skills interventions on learners and teachers for the EPPI-Centre's Research Evidence in Education Library (REEL) commissioned by the Department for Education and Skills. It examines the benefits of systematic reviewing for reinforcing the 'evidence base' and its limitations in terms of providing guidance for new developments.

(4866)

In this paper we are trying to reflect more broadly on the implications of our work in a very specific field: the use of thinking skills interventions with pupils in mainstream schooling. The focus of this work has been a relatively small empirical problem, which we have broken down into a series of key questions. Do thinking skills interventions ‘work’? In what ways do they work, with which groups and for how long? Are they different in a significant way from other kinds of pedagogical interventions? Over the last 10 years the Thinking Skills Research group at the University of Newcastle upon Tyne have been looking at these questions through developmental and experimental work with practitioners, supported by a series of narrative reviews (see, for example Baumfield and Oberski, 1998; Higgins and Leat, 1997; Leat and Higgins, 2002).

In the course of our most recent review of Thinking Skills for the EPPI-Centre the research team have amassed a database of almost 900 articles, papers and books. We have undertaken three reviews: on the impact of interventions in classrooms; on the impact on teachers and a meta-analysis on the impact on pupil attainment. In this paper we want to look critically at what the systematic review and meta-analysis processes have revealed about thinking skills interventions and what they have not revealed. We intend to problematise issues of comparability and homogeneity in this kind of research and highlight the dangers of ‘blending’ results together to produce headline results. We will present results which indicate where there is strong evidence for the value of thinking skills – particularly in relation to cognitive and curricular gains - and explore where evidence is less coherent. For thinking skills interventions, it appears that the effects on learners are age and context specific, so that, picking up the analogy in the title, we need to re-examine the nature of the strawberry, rather than the mixed fruit jam that can result from meta-analysis. However, we want also to argue that the tightly focussed, ‘funnelled’ results of very specific analytic reviews need to be balanced against the broader perspectives of narrative review.

Underlying these issues there remains the problem of the kinds of research papers which are being published – the silences in data reporting and the formats in which journals accept empirical work. There are methodological tensions between the criteria of systematic reviewing with the close focus required in generating research questions in these reviews and the holistic, enthusiastic approach of many reports which focus on development work with teachers and learners. Trends in the United States towards a research design that is ‘oven-ready’ for meta analysis are significantly limiting the kinds of information available about learning and teaching. Moreover, as Slavin (2004) remarks, the policy discourse of a particular kind of evidence based research is running ahead of the educational research paradigm: put simply, there are not very many studies which fit the criteria set by bodies like the Clearinghouse (www.w-w-c.org) and many of these, rather than reporting impacts in authentic classroom interventions are classified by Slavin as ‘wierdo lab studies’ with small groups, short durations and un-natural contexts. We argue that until we can achieve recognition of and consensus about the strengths, weaknesses and contribution of the various methods current in educational research, we will continue to alternately reify and downplay methods, creating artificial opposition and weakening our ability to convey meaningful messages to practitioners. The process of literature review is not, as the hyperbole might suggest, to produce definitive answers in a given field but is essentially one of producing better questions for the

next round of empirical work by academics and practitioners. There is a strong case for more triangulation of review techniques, drawing on the compensatory strengths of each approach to produce a view of the field that has both ‘headlines’ of greatest effects, rigour and transparency in the selection of studies and depth and nuance in the interpretation of results.

The traditional approach to literature review in educational research has been a narrative one and has been extensively criticised as unsystematic in defining or covering the chosen field, over-sensitive to publication bias and because of the lack of explicit criteria for inclusion and exclusion, open to the accusation of being subjective or superficial (see, for example, Hunter and Schmidt, 1990). The synthesis of results and the weighting of evidence are dependent on the prejudices and preferences of the reviewer, with no clarity about study quality. In addition, quantitative results in this form of review are generally reported in terms of p values, rather than in effect sizes (Fan, 2001). It should be noted that the majority of these criticisms, valid though they are, have been expressed by advocates of systematic review and meta-analysis. Narrative approaches to literature reviews do have several counterbalancing advantages: they are sensitive to context and can provide thematic depth in a way denied to systematic reviews by the rigidity of their inclusion criteria; they can give appropriate weight to exemplary studies and they can provide context-rich information about educational innovations in a way that is accessible to practitioners.

Systematic reviews are in some ways an improvement on narrative reviews because the transparency of the search and selection strategies allows the reader to assess the relevance of the review to her own concerns and to make a judgement about how comprehensive the reviewers have been. The promise of a systematic review is that the selected area of focus will have been thoroughly examined and the studies within it subject to the same rigorous grading procedure, so that conclusions drawn from them will be based on ‘best evidence synthesis’ (Slavin, 1995). However, the process of systematic review is not without controversy: there is a view that the decision process about the criteria limiting the pool of studies in the review should ideally be iterative, drawing on the current understandings in the field and the studies available. This is opposed by those who feel that an iterative process renders the review more impressionistic but in turn pre-set criteria are criticised as more prone to producing an artificially limited view of the field. The much-vaunted advantages of the systematic review process, which uses the inclusion criteria to funnel studies, producing a ‘distilled truth’ has to be set against the restrictions of the ‘view through the funnel’ when attempting to look at terrain as diverse and context dependent as education.

Olkin (1992) reminds us that “doing a meta analysis is easy, doing one well is hard” and this is often attributed to the three key ‘validity threats’ (Sharpe, 1997). Firstly, the unavailability of negative or inconclusive studies, also known as the ‘file drawer’ phenomenon; secondly, striking the balance between ‘selection bias’ – a key criticism of narrative review– or including too many poor quality studies which skew the overall effect size and thirdly, the ‘apples and oranges’ problem of deciding which studies, research populations and interventions are genuinely comparable. Meta analysis can give clear indications of ‘what works’ by producing comprehensible information from the complexities of multiple studies and populations (Lipsey and Wilson, 2001). Olkin might further have remarked that doing a meta analysis is tempting, since it offers, through the comfort of numbers and the promise of

impartiality, the possibility of ‘definitive’ answers to research questions. Most commentators, while resisting this tempting mirage, find it impossible to avoid claiming that meta analysis has greater impartiality, validity and explicative power, ignoring the problem that the synthesis of multiple studies reflects the general direction of a trend, rather than an explanation of that trend.

One way of moving beyond description to explanation (Cook et al 1992), is to recombine the methodological approaches to reviewing, since it becomes clear that ‘good quality’ reviews from all three approaches share key concerns about validity (Hunter and Schmidt, 1990; Slavin, 1995; Sharpe; 1997). Different key questions related to the size of the effect, the nature of the effect in specific contexts and the variation of effects across contexts can be answered by meta analysis, systematic review and narrative review, respectively.

Thinking Skills Interventions

The teaching of thinking skills is an explicit part of the National Curriculum in England and Wales and contributes directly to the DfES’s current initiative ‘Teaching and Learning in the Foundation Subjects’ at Key Stage 3. The descriptive review by Carol McGuinness (1999) provided an overview of current research into the teaching of thinking skills and built on the work of earlier reviews in this area. Nisbet and Davies (1990) listed 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 amongst 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 have been in existence for somewhat longer, but the term itself is ambiguous and there is disagreement about how it relates to aspects of pedagogy more broadly. Our working definition for the purposes of this series of reviews is that thinking skills interventions can be identified as 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 characterized as approaches or programmes which require learners to articulate and evaluate specific learning approaches and which identify specific cognitive, 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 consisted of six major categories arranged in the following order *Knowledge, Comprehension, Application, Analysis, Synthesis* and *Evaluation*.

Examples of programmes and approaches commonly used in schools are Instrumental Enrichment (Feuerstein, Rand, Hoffman, and Miller, 1980), Philosophy For Children (Lipman, Sharp and Oscanyan, 1980) Cognitive Acceleration Through Science Education (Adey, Shayer and Yates, 1995), or Somerset Thinking Skills (Blagg, Ballinger and Gardner, 1988). Considerable interest has also been shown in how these

more formal programmes can be integrated effectively or ‘infused’ into teaching approaches and adopted more widely by teachers (McGuinness, Wylie, Greer, & Sheehy, 1995; McGuinness, 1999; Leat and Higgins, 2002). Our interpretation of the literature is that an in-depth analysis is needed to evaluate the claims of any impact of such approaches on teaching and learning in classrooms. A further reasonable aim is therefore to try to identify any common features of the impact of implementing thinking skills approaches and to consider how well these relate to wider findings about teaching and learning more broadly, such as formative assessment and feedback in classrooms (Torrance and Pryor, 1998; Black and Wiliam, 1998) or classroom talk and interaction (Mercer, 1995; Galton, 1999). In the reviews carried out by our group, it has been our intention to maintain this focus on the relationship between the general characteristics of thinking skills approaches and interventions and any impact on teaching and learning.

Thinking skills approaches are generally popular with teachers and there is evidence that they seem to support changing patterns of interaction in classrooms (Baumfield and Oberski, 1998; 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, 1994), ‘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) and collaborative group work (Webb and Farrivar, 1994; Galton et al., 1999).

The meta-analysis: what we did and what we meant by it.

The review was conducted as one of the EPPI-Centre¹ co-ordinated reviews to develop an evidence base to inform policy and practice funded by the Department for Education and Skills. The review group set out to answer a number of questions over the three years of funding for the reviews. These questions were clustered around our central question about what studies were available and relevant to answering the broad question ‘What is the impact of the implementation of thinking skills interventions on teaching and learning?’ The meta-analysis is the third phase of this systematic review. The first phase addressed the question of evidence for impact on learners and the second turned to the evidence for impact on teachers.

In our first review we identified and described 191 studies up until 2002. We used narrative synthesis methods to address the question ‘what is the impact of thinking skills interventions on pupils?’. Twenty-three studies were included and reviewed in depth (Higgins *et al.* 2004). This review identified a number of positive findings about thinking skills approaches but concluded that the selection and implementation of thinking skills approaches as means to improve teaching and learning 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.

¹ <http://eppi.ioe.ac.uk/EPPIWeb/home.aspx>

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?” (e.g. Marzano *et al.* 2001, Hattie *et al.* 1996).

In our meta-analysis we tested the hypothesis that thinking skills interventions have a positive impact by addressing the following questions:

1. What is the magnitude of the quantitative impact of thinking skills interventions on pupils’ cognitive achievement?
2. What is the magnitude of the quantitative impact of thinking skills interventions on pupils’ curriculum attainment?
3. What is the magnitude of the quantitative impact of thinking skills interventions have on pupils’ affective states?

It was our expectation that intervention effects would be positive, since that vast majority of educational innovations have a positive effect but we were interested in whether they would achieve above the 0.4 level cited by Hattie as the average intervention effect size from his meta-analysis of 200,000 effect sizes. Indeed, given that there are interventions with effects ranging from 0.6 to over 1, he considers 0.5 as a minimum for an intervention to be considered ‘educationally significant’ (Hattie, 2004).

Method

For the first review the team had conducted a broad sweep of the education literature by applying a range of relevant keywords to 16 databases (such as British Education Index, ERIC, PsychINFO, Education Abstracts, etc.) using the online gateways of BIDS, Web of Science and First Search. This yielded references to almost 6,500 books, chapters, dissertations and academic articles which formed the basis of the field. An analysis of 1500 of these items indicated that the majority of references (61%) were concerned with pupils’ thinking in school settings as opposed to further and higher education or the use of thinking skills approaches in the workplace. The focus of the review required sources which, implicitly or explicitly, sought to evaluate the implementation of thinking skills programmes in the classroom and so the team identified a subset of 896 by screening abstracts and titles which were subjected to a detailed analysis to fit the review requirements. Criticisms relating to the quality of educational research are not new (e.g. Hillage *et al.*, 1998; Tooley and Darby, 1998) but we feel it is necessary to comment on the quality of abstracting we encountered. Descriptions were often misleading, frequently over-played the amount or types of data actually reported and displayed a tendency to report findings which were inferences rather than conclusions from the data. The screening of abstracts often therefore proved inconclusive and we needed the full paper to determine whether or not the study met the criteria.

At this stage studies were included which:

1. Were set in a school or schools and are concerned with any section of the school population (including pupils with special needs).
2. Evaluated the impact of the implementation of thinking skills interventions on teaching and/or learning.
3. Were concerned with the phases of compulsory schooling (5 –16 year olds).

4. Contained empirical classroom research with data or evidence (pupil outcomes, classroom processes, teacher's views).

Thinking skills interventions were 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 learning and/or teaching. These interventions could be taught as separate programmes or infused into curriculum teaching. Impact included, for example, pupil and teacher motivation and engagement; patterns of classroom interaction; self-regulation and metacognitive monitoring as well as learners' attainment. 191 studies met all these specifications and were used by the team to form a descriptive 'map' of the field. Of these around 100 appeared to contain quantitative data about impact on pupil attainment.

For the meta-analysis additional inclusion criteria needed to be applied to these studies which examined the impact of thinking skills interventions on pupils. To be considered, studies needed to contain quantitative data of the impact of thinking skills approaches on pupils. This quantitative data had to include either a reported effect size or sufficient detail to calculate an effect size for at least one of the impact measures used. This included standardised tests of cognitive ability (such as Raven's Progressive Matrices); standardised tests of curriculum attainment (such as in reading or mathematics); criterion referenced tests (such as a school science test); tests of impact on affect or conation (such as attitude or self-efficacy scales). The studies also needed to use a control or comparison group which had not received the intervention. We also rejected studies with fewer than 10 research subjects. The initial search was also repeated to identify any studies which met both sets of criteria that had been published since 2002). When these, as well as a small number which focused on children with severe or complex special educational needs, were removed there were 30 studies with evidence about the impact of thinking skills approaches on which to perform the meta-analysis (see Table 1 below). All 30 of these studies were published, 27 as peer-reviewed journal articles, three as chapters in edited collections.

These 30 studies have been reviewed in depth according to the data extraction guidelines for the EPPI-Centre reviewing process (EPPI, 2001), two studies contained effect sizes but insufficient information to calculate the confidence interval and therefore combine the findings with the other studies; and one study contained effect sizes but we have been unable to confirm the calculations using other data presented in the report (we have asked one of the authors of the paper to check the figures from their original data). The data from the studies was initially converted to Cohen's *d* using Coe's (2003) Effect Size calculator.

Table 1: Characteristics of the studies in the meta-analysis

Study	Thinking skills programme or approach	School phase	Intervention effects measured	Effect sizes <i>d</i>
Adey (1990)	CASE	Secondary	Cognitive (PRT)	0.21
Adey (2002)	CASE	Primary	Cognitive (PRT)	0.43, 0.47
Cardelle-Elawar (1992)	Metacognitive	Primary	Cognitive (Ravens) Curriculum (Maths) Curriculum (Basic Skills) Affective (attitude to maths)	1.11, 1.42 5.93, 6.56 0.2, 0.74 8.55, 5.36
Chang (1999)	Search, Solve, Create and Share	Secondary	Cognitive Curriculum (Science)	0.7 0.3
Collings (1994)	Formal operations,	Secondary	Cognitive	1.09, 1.13, 1.53

	metacognition		Curriculum (Science)	0.36, 0.68, 0.92
Csapo (1992)	Operational abilities	Primary and Secondary	Cognitive (logical operations) (combinative operations) (systematizing operations)	P= 0.4, 0.64, 0.48, S=0, 0.05, 0.01 P=0.91, 0.8, 0.39, S= 0.4, 0.32, 0.09 P=-0.48, -0.29, 0.08, S=-0.16, 0.09, -0.1
Cunningham (2002)	Skills for Positive thinking	Primary	Affective	0.23; -0.22; 0.01; 0.38; 0.2
De Koning (1999)	Inductive reasoning	Primary	Cognitive (Ravens) Curriculum (Reading)	0.84 1.24
Donegan (1998)	Vernon	Primary	Affective	0.65
Greenberg (2000)	COGNET/ CEA	Secondary	Curriculum (Maths) Curriculum (Reading)	1, 1.5, 1.1 0.55, 1, 0.2, 0.9, 0.3, 0.9
Haywood (1998)	Feuerstein Enrichment	Secondary	Cognitive (Ravens) Cognitive (PMA) Curriculum (Maths) Curriculum (Reading)	0.7 1.21, 0.77 -0.46, -0.54, -0.33, -0.18, -0.62 -0.18
Hoek (1999)	Metacognitive	Secondary	Curriculum	0.2, 0, 0.38
Iqbal (2000)	CASE	Secondary	Curriculum (Maths) Curriculum (Science)	0.004, 0.3, -0.35, 0.43 0.07, -0.7, 0.16, 0.16, 0.67, 0.87
Kaniel (1992)	Feuerstein's Instrumental Enrichment	Secondary	Cognitive (Ravens) Cognitive (Analogies and organiser)	0.67 0.41, 0.83, 1.53
Kramarski (1997)	Metacognitive	Secondary	Cognitive Curriculum (Maths)	0.19, 0.47, 1.08, 2.90, 3.28, 4.44 0.39
Maqsud (1998)	Metacognitive	Secondary	Cognitive Curriculum (Maths) Affective	0.98, 0.83 0.86, 1.19, 1.80, 1.22, 1.21 0.73
Martin (1984)	Feuerstein's Instrumental Enrichment	Secondary	Cognitive (Ravens) Curriculum (Maths)	0.89 0.38
Mercer (1999)	Talk, Reasoning and Computers	Primary	Cognitive (Ravens)	0.32
Muttart (1984)	Feuerstein's Instrumental Enrichment	Secondary	Cognitive (PMA) Curriculum (achievement) Affective	0.89, 0.97 -0.5, 0.41, 0.02, 1.17
Naval-Severino (1993)	Creative thinking	Primary	Cognitive (Torrance TCT)	1.6, 1.57, 1.38
Oladunni (1998)	Metacognitive/ Heuristic	Secondary	Cognitive (creative maths)	1.04
Riding (1987)	Critical Thinking	Primary	Cognitive (Ravens) Curriculum (Maths) Curriculum (Reading)	-0.05, 0.79, 0.68, 1.42 -0.82, -0.77, -0.49, 0.69 -0.17, 0.06, 1.10, 1.32
Ritchie (1996)	CoRT (De Bono)	Secondary	Cognitive (CoRT)	0, 0.19, 0.32
Scheinin (1999)	Cognitive education	Secondary	Cognitive Affective	-0.04; 0.3; 0.13; 0.28; 0.43, 0.42; 0.57 0.02; 0.6; 0.11; 0.78, -0.08
Schmid (1990)	Concept mapping	Secondary	Curriculum (Reading)	0.23, 0.40, 0.43
Shayer (1987)	Feuerstein's Instrumental Enrichment	Secondary	Cognitive Curriculum (Maths)	1.07-1.22 0.21; 0.37; 0; 0.23; 0.98; -0.26; -0.35; 0.07; 0.36; 0.26; 0.47; 0.21; 0.57; 0.46
Strang (1993)	Feuerstein Enrichment	Secondary	Curriculum (Science)	1.15
Tenenbaum (1986)	CPR+FB/C	Secondary	Curriculum (Maths) Curriculum (Science)	1.71 1.75
Tzurriel (1994)	Feuerstein's Instrumental Enrichment	Secondary	Cognitive (FE)	0.21
Ward (1993)	Think-Aloud	Primary	Curriculum (Reading)	1.02, 1.61

Sample sizes were in the range 10->900 and because of this we conducted tests which confirmed that sample size and effect size² were not significantly related using a 'funnel' plot (Lipsey and Wilson, 2001): see figure 1 below.

² Effect sizes were entered into the EPPI-Centre database and analysed using the on-line meta-analysis tools. This database uses Hedges g as the standard effect size.

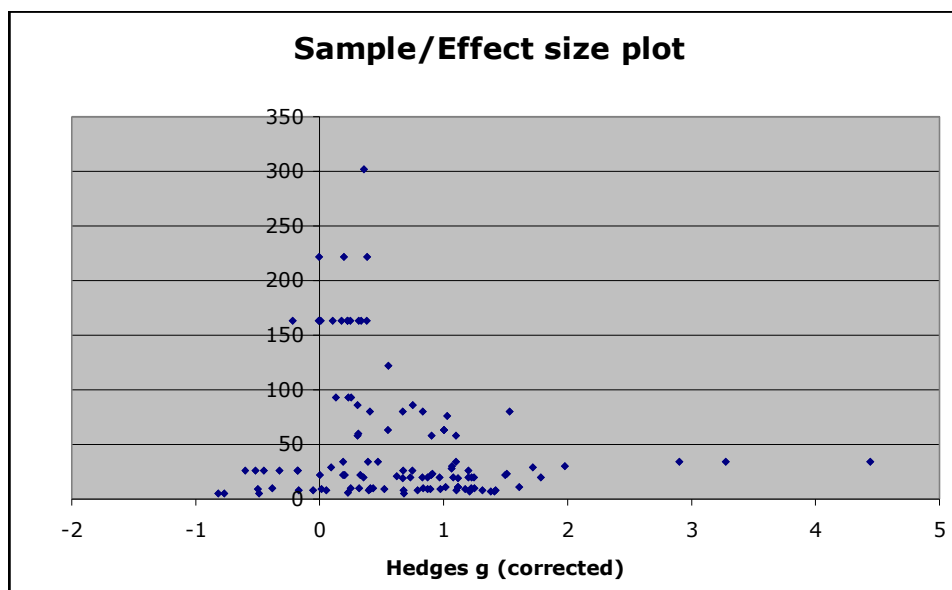


Figure 1 Effect size plotted against sample size

Results of the meta analysis

Analysis of these studies indicate that thinking skills approaches are effective in improving pupils' learning. A meta-analysis of this impact found an overall effect size of 0.71 on cognitive measures (such as tests of reasoning or non-verbal measures such as Ravens Progressive Matrices) and an effect size of 0.66 for curriculum outcomes (such as 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 on curriculum tests and to about 24th place on cognitive measures.

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.

A breakdown of the effect sizes and types of outcomes is presented in Table 2. What is interesting is that although the average impact of thinking skills interventions on all outcome measures was similar to the overall average effect size of all interventions reported by Hattie (2004), the impact on cognitive measures (the key outcome for such interventions) was considerably higher (0.71). Furthermore attempts to improve pupils' thinking do seem to have a positive effect on their attainment in curriculum subjects (0.66). Study heterogeneity was calculated by the meta-analysis software in the EPPI-Reviewer database.

Table 2: Effect sizes and outcome measures

Type of outcome	ES	CI	No of effects/No of studies
All outcomes	0.47	0.43 - 0.50	119 effects from 27 studies ³
Cognitive outcomes	0.71	0.63 - 0.78	35 effects from 15 studies ⁴

³ Heterogeneity statistic $Q = 1.1E+03$ $df = 120$ $p = 0$ Test statistic (combined effect) $z = 23.6$ $p < 0.001$ Inverse Variance (fixed effects model)

⁴ Heterogeneity statistic $Q = 195$ $df = 34$ $p = 0$ Test statistic (combined effect) $z = 17.9$ $p < 0.001$ Inverse Variance (fixed effects model)

Curricular outcomes	0.66	0.59 0.74	55 effects from 17 studies ⁵
Affective outcomes	0.26	0.20 0.33	22 effects from 5 studies ⁶
Creative outcomes	0.51	0.36 0.65	8 effects from 4 studies ⁷

Meta-analysis does appear to be finding consistent messages in the educational research literature. Our study found an overall mean effect of 0.47, similar to that of Hattie's vast database of meta-analyses (Hattie, 2004). Looking at a smaller part of our study, the impact of Feuerstein's Instrumental Enrichment which is one of the most extensively researched of the thinking skills programmes, our results broadly concur with that of Romney and Samuels' (2001) study which found moderate overall effects and an effect size of 0.43 on reasoning ability (p.28) – our findings were similar with an overall effect size of 0.32 (26 effects from 5 studies) and an effect size of 0.49 on tests of reasoning (6 effects from 4 studies). This 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 etc., but that this works because of the magnitude of impact.
- We need to be building a model based on these relative magnitudes of effects.”

Of course interpreting the model that and the relative magnitude of effects will always be challenging. The messages may well not be consistent across different contexts. This in turn suggest 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) point out, "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

⁵ Heterogeneity statistic $Q = 443$ $df = 55$ $p = 0$ Test statistic (combined effect) $z = 17.6$ $p < 0.001$ Inverse Variance (fixed effects model)

⁶ Heterogeneity statistic $Q = 364$ $df = 21$ $p = 0$ Test statistic (combined effect) $z = 7.73$ $p < 0.001$ Inverse Variance (fixed effects model)

⁷ Heterogeneity statistic $Q = 21.7$ $df = 7$ $p = 0.00282$ Test statistic (combined effect) $z = 6.77$ $p < 0.001$ Inverse Variance (fixed effects model)

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 identification of 'thinking skills' 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'. The implications are perhaps clearer at policy level. There are already signs that the Department for Education and Skills have 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.

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) or 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 paper adds weight to this growing body of evidence. However the variation in the impact of thinking skills approaches by age or gender (e.g. Csapó 1992; Adey and Shayer, 1993) combined with the differences between the programmes themselves make it difficult to offer more than Bassey's Best Estimate of Trustworthiness. Meta-analysis and the rigours of systematic reviewing give us a degree of confidence that the evidence we cite is reliable but without the contextual sensitivity and detail which characterises a traditional narrative approach to research literature, we cannot make meaningful recommendations to practitioners.

Hence the title of our paper. The boiling down of information through the systematic review and meta-analysis process produces the illusion that the data itself is homogenised, that the various intellectual fruits have made an undifferentiated jam. However, the fruit jam that represents the meta-analysis is not what a practitioner needs. They need to act in a particular context and to make a particular choice about which approach is likely to work for them in their school – or to pick a nice fresh strawberry out of the jam.

References

- 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 & Instruction*, 11(1), pp. 1-29.
- Adey, P.S., Shayer, M. and Yates, C. (1995) *Thinking Science: The Curriculum Materials of the CASE Project* London: Thomas Nelson and Sons.
- Bassey, M. (2000) *Fuzzy Generalisations And Best Estimates Of Trustworthiness: a step towards transforming research knowledge about learning into effective teaching practice*. TLRP Annual Conference 2000.
<http://www.tlrp.org/acadpub/Bassey2000.pdf>
- Baumfield, V.M. and Oberski, I.O (1998) What do Teachers Think about Thinking Skills? *Quality Assurance in Education*, 6:1, pp 44-51.
- Black, P. and Wiliam, D. (1998) Assessment and Classroom Learning *Assessment in Education* 5.1. pp 7-71
- Blagg, N., Ballinger, M. and Gardner, R. (1988) *Somerset Thinking Skills Course Handbook* Oxford: Basil Blackwell.
- Bloom, B. S. (Ed.) (1956) *Taxonomy of Educational Objectives, the classification of educational goals - Handbook I: Cognitive Domain*. New York: McKay
- Csapó, B. (1992) Improving Operational Abilities in Children In A. Demetriou, M. Shayer and A. Efklides (Eds) *Neo-Piagetian Theories of Cognitive Development: Implications and Applications For Education* London: Routledge.
- Dweck, C. (1999) *Self Theories: Their Role in Motivation, Personality and Development*. Hove (Sussex): Psychology Press (Taylor & Francis).
- Edwards, A. and Westgate, D. (1987) *Investigating Classroom Talk*. London: Falmer Press.
- EPPI (2001) *EPPI-Centre Review Group Manual Version 1.1*, London: Social Science Research Unit, Institute of Education, University of London.
- EPPI (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.
- Fan, X. (2001) *Statistical significance and effect size in educational research: two sides of a coin*. *Journal of Educational Research*, 95 (5) 275-282
- Feuerstein, R. Rand, Y., Hoffman, M.B. and Miller, R. (1980) *Instrumental Enrichment: an intervention programme for cognitive modifiability* Baltimore: University Park Press.
- Galton, G., Hargreaves, L., Comber, C. and Pell, A. (1999) *Inside the Primary Classroom 20 Years On*, London: Routledge.
- Hamers, J.H.M. & Overtoom, M.Th. (Eds.). (1997). *Teaching thinking in Europe. Inventory of European programmes*. Utrecht: SARDES
- Hattie, J., Biggs, J., and Purdie, N. (1996). *Effects of learning skills interventions on student learning: A meta-analysis*. *Review of Educational Research*, 66, 99-136.
- Hattie, J. (1999) "Influences On Student Learning". Inaugural Lecture: Professor of Education. University of Auckland. August 2, 1999.
<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, July 8th, 2004
- Higgins, S. (2001) *Thinking Through Primary Teaching*. Cambridge, Chris Kington

Publishing.

Higgins, S. and Leat, D. (1997) "Horses for courses or courses for horses: what is effective teacher development?" *British Journal of In-Service Education* 23.3 pp 303-314

Higgins, S. and Moseley D. (2002) 'Raising Achievement in Literacy through ICT' in M. Monteith (ed), *ICT and Literacy* Buckingham: Open University Press.

Higgins S, Baumfield V, Lin M, Moseley D, Butterworth M, Downey G, Gregson M, Oberski I, Rockett M and Thacker D (forthcoming) *Thinking skills approaches to effective teaching and learning*. In: Research Evidence in Education Library. London: EPPI-Centre, Social Science Research Unit, Institute of Education.

Hillage, J., Pearson, R., Anderson, A. and Tamkin, P. (1998). *Excellence in research on schools*. London: DfEE.

Hunter, J. E. and Schmidt, F. L. (1990) *Methods of Meta Analysis: correcting error and bias in research findings* New York: Sage

Leat, D. and Higgins S. (2002) The role of powerful pedagogical strategies in curriculum development *The Curriculum Journal* 13.1 pp 71-85 ISSN 0958-5176.

Leat, D. and Lin, M. (2003) Developing a Pedagogy of Metacognition and Transfer: some signposts for the generation and use of knowledge and the creation of research partnerships *British Educational Research Journal* 29.3 pp 383 – 414.

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

Lipsey, M. W., and Wilson, D. B. (2001). *Practical Meta-Analysis. Applied Social Research Methods Series* (Vol. 49). Thousand Oaks, CA: SAGE Publications

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, N.A.F. (1995) Developing children's thinking: A tale of three projects, *Irish Journal of Psychology*, 16(4), pp. 378-388.

Marzano, R. J., Pickering, D. J., & Pollock, J. E. (2001). *Classroom instruction that works: Research-based strategies for increasing student achievement*. Alexandria, VA: Association for Supervision and Curriculum Development.

Mercer, N. (1995) *The guided construction of knowledge: Talk amongst teachers and learners* Clevedon: Multilingual Matters.

Nisbet, J. and Davies, P. (1990) The curriculum redefined: learning to think-thinking to learn, *Research Papers in Education*, 5.1, pp. 49-92.

Olkin, I. (1992) *Meta-analysis: methods for combining independent studies*. *Statistical Science*, 7, 226

Romney, D.M. and Samuels, M.T. (2001) A meta-analytic evaluation of Feuerstein's Instrumental Enrichment program *Educational and Child Psychology* 18(4) pp.19-34

Slavin, R. (2004) *Evidence-Based Education Policies: Promises and Pitfalls* Keynote Address, British Educational Research Association Conference, UMIST, Manchester, September 15-18th September.

Tooley, J. and Darby, D. (1998) *Educational research - a critique*. London: Office for Standards in Education

Torrance, H. and Pryor, J. (1998) *Investigating formative assessment: teaching, learning and assessment in the classroom* Open University Press.

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

Webb, N. and Farivar, S. (1994) Promoting helping behaviour in cooperative small groups in middle school mathematics. *American Educational Research Journal*, 31.2 pp. 369-396.

Wood, D. and Wood, H. (1996) Vygotsky, Tutoring and Learning, *Oxford Review of Education*. Vol. 22.1, pp. 5-16.

Studies in the meta-analysis

1. Adey, P. and M. Shayer (1990). "Accelerating the Development of Formal Thinking in Middle and High-School-Students." *Journal of Research in Science Teaching* 27(3): 267-285.
2. Adey, P., A. Robertson, et al. (2002). "Effects of a cognitive acceleration programme on Year I pupils." *British Journal of Educational Psychology* 72: 1-25.
3. Cardelle-Elawar, M. (1992). "Effects of Teaching Metacognitive Skills to Students with Low Mathematics Ability." *Teaching & Teacher Education* 8(2): 109-21.
4. Chang, C.-Y. and J. P. Barufaldi (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(4): 373-88.
5. Collings, J. N. (1994). "Some fundamental questions about scientific thinking." *Research in Science & Technological Education* 12(2): 161-173.
6. Csapó, B. (1992). Improving operational abilities in children In A. Demetriou, M. Shayer and A. Efklides. *Neo-Piagetian theories of cognitive development* London, Routledge.
7. Cunningham, E.G., Brandon, C.M. and Frydenberg, E. (2002) Enhancing coping resources in early adolescence through a school-based program teaching optimistic thinking skills. *Anxiety Coping and Stress* 15(4): 369-381
8. De Koning, E. and J. H. M. Hamers (1999). Teaching inductive reasoning: theoretical background and educational implications. In J. H. M. Hamers, J. E. H. Van Luit and B. Csapó. *Teaching and Learning Thinking Skills* Abingdon, Swets and Zeitlinger.
9. Donegan, A. L. and J. O. Rust (1998). "Rational Emotive Education for Improving Self-Concept in Second-Grade Students." *Journal of Humanistic Education & Development* 36(4): 248-56
10. Greenberg, K. H. (2000). "Attending to hidden needs: The Cognitive Enrichment Advantage perspective." *Educational & Child Psychology*. Special Issue: Psychological influences upon educational intervention 17(3): 51-69.
11. Haywood, H. C. and et al. (1988). "Cognitive Education with Deaf Adolescents: Effects of Instrumental Enrichment." *Topics in Language Disorders* 8(4): 23-40.
12. Hoek, D., P. van den Eeden, et al. (1999). "The Effects of Social and Cognitive Strategies Instruction on the Mathematics Achievement in Secondary Education." *Learning & Instruction* 9: 427-448.

13. Iqbal, H. M. and M. Shayer (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(3): 259-274.
14. Kaniel, S. and R. Reichenberg (1992). "Instrumental Enrichment--Effects of Generalization and Durability with Talented Adolescents." *Gifted Education International* 8(3): 128-35
15. Kramarski, B. and Z. R. Mevarech (1997). "Cognitive-metacognitive training within a problem-solving based Logo environment." *British Journal of Educational Psychology* 67(4): 425-445
16. Maqsd, M. (1998). "Effects of metacognitive instruction on mathematics achievement and attitude towards mathematics of low mathematics achievers." *Educational Research* 40(2): 237-243.
17. Martin, D. S. (1984). "Cognitive Modification for the Hearing Impaired Adolescent: The Promise." *Exceptional Children* 51(3): 235-42.
18. Mercer, N., R. Wegerif, et al. (1999). "Children's talk and the development of reasoning in the classroom." *British Educational Research Journal* 25(1): 95-111.
19. Muttart, K. (1984). "Assessment of Effects of Instrumental Enrichment Cognitive Training." *Special Education in Canada* 58(3): 106-08.
20. Naval-Severino, T. (1993). "Developing creative thinking among intellectually able Filipino children from disadvantaged urban communities." *Gifted Education International* 9(2): 119-123.
21. Oladunni, M. O. (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 & Technology* 29(6): 867-874.
22. Riding, R. J. and S. D. Powell (1987). "The Effect on Reasoning, Reading and Number Performance of Computer presented Critical Thinking Activities in Five-year-old Children." *Educational Psychology* 7(1): 55-65.
23. Ritchie, S. M. and J. Edwards (1996). "Creative Thinking Instruction for Aboriginal Children." *Learning & Instruction* 6(1): 59-75.
24. Scheinin, P. M. and J. Mehtäläinen (1999). Applying the theory of knowledge to teaching thinking. In J. H. M. Hamers, J. E. H. Van Luit and B. Csapó. *Teaching and Learning Thinking Skills* Abingdon, Swets and Zeitlinger.
25. Schmid, R. F. and G. Telaro (1990). "Concept Mapping as an Instructional Strategy for High School Biology." *Journal of Educational Research* 84(2): 78-85.
26. Shayer, M. and F. Beasley (1987). "Does instrumental enrichment work?" *British Educational Research Journal* 13(2): 101-119.

27. Strang, J. and M. Shayer (1993). "Enhancing high school students' achievement in chemistry through a thinking skills approach." *International Journal of Science Education* 15(3): 319-337.
28. 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(2): 105-114.
29. Tzuriel, D. and M. Alfassi (1994). "Cognitive and Motivational Modifiability as a Function of the Instrumental Enrichment (IE) Program." *Special Services in the Schools* 8(2): 91-128.
30. Ward, L. and D. Traweek (1993). "Application of a Metacognitive Strategy to Assessment, Intervention, and Consultation: A Think-Aloud Technique." *Journal of School Psychology* 31(4): 469-85.