BERA
BRITISH EDUCATIONAL RESEARCH ASSOCIATION

Keynote symposium
Informing Policy and Influencing Practice: Messages from Cross-Disciplinary Research on Achievement, Engagement and Participation in Mathematics and Science

Research-informed pedagogical innovation at scale in school mathematics and science education

Kenneth Ruthven
University of Cambridge

In the first part of this presentation I will look back at the recent history of research-informed pedagogical innovation at scale in the primary and lower-secondary phases of schooling in England. I will then examine the evolving state of student attainment and attitude in these subjects at the end of lower-secondary schooling, drawing in particular on evidence from the TIMSS and PISA international study series.

In the second part of the presentation I will look forwards to consider future needs for research-informed innovation at scale in the light of evolving knowledge about effective pedagogy in school mathematics and science. I will sketch three types of recent systematic synthesis of research in this area and examine what guidance they offer for efforts to support pedagogical improvement at scale, such as the *epISTEMe* project which I am currently leading.

[Note: this paper is an updated summary of one (Ruthven, 2011) published in the recent special issue of the *International Journal of Science and Mathematics Education* on Enhancing the participation, engagement and achievement of young people in science and mathematics education. This updating has taken account of recent PISA evidence about attainment, and provides broader consideration of the TIMSS attitude constructs. The summarisation has aimed to outline the argument for certain key types of pedagogical development being crucial for improvement at scale in the teaching of mathematics and science In English schools.]

The National Strategies: Pedagogical experiment at systemic scale

- A systemic school improvement programme
  - Introduced to primary schools from 1998 (informally) and 1999 (formally)
  - Extended to lower secondary for student cohorts entering 2001 (mathematics) & 2002 (science)
- Independent evidence of impact provided by regular international study series such as TIMSS and PISA
  - Data gathered about attitude as well as attainment
  - Comparison across systems and between subjects

Between 1998 and 1999, the National Strategies launched a systemic improvement programme for primary schooling in England, extended to the lower-secondary phase between 2001 and 2002. The scale of the programme and the priority given to it were such that it was by far the dominant influence on school mathematics and science teaching at these levels over the ensuing decade.

At regular intervals (since 1995 in the case of TIMSS and 2000 in the case of PISA) international study series have collected cross-system evidence about the outcomes of compulsory education. This evidence concerns students' attitudes as well as their levels of attainment. The use of common methods and instruments provides scope for comparison both across systems and between subjects. In view of their dominant position, it is highly plausible that trends in these study series reflect the impact of the National Strategies, showing any resultant changes in student outcomes in mathematics and science.

The National Strategies: The influence of pedagogical research

- Main basis was a predominantly American body of “process-product” research on effective teaching
- Core model of “active teaching” had been validated in relation to basic mathematical knowledge and skills
- Other research suggested that “additional classroom processes… needed to enhance higher order thinking:
  - a focus on meaning and understanding…,
  - direct teaching of higher level cognitive strategies and problem-solving,…
  - co-operative small group work.”

(Reynolds & Muijs, 1999, p. 281)

The process through which the pedagogical model promoted by the National Strategies was developed was highly politicised. In particular, the preconceptions of policy-makers resonated more strongly with the perspectives of “school improvers” than with those of “mathematics educators”, while“science educators” were marginalised at this stage.

Consequently, the main influence on the pedagogical model recommended for mathematics (which subsequently shaped similar recommendations for science) was a predominantly American body of “process-product” research on effective teaching. The core model of “active teaching” had been developed and validated primarily through research on basic instruction in mathematical knowledge and skill during the 1970s and early 1980s (Good, Grouws, & Ebmeier, 1983; Reynolds & Muijs, 1999).

Nevertheless, the chair of the Task Force that devised the pedagogical model guiding the Numeracy Strategy acknowledged that, to successfully develop higher-order thinking in mathematics, this “interactive teaching” needed to be complemented by:

> “a focus on meaning and understanding…, direct teaching of higher level cognitive strategies and problem-solving,… co-operative small group work.” (Reynolds & Muijs, 1999, p. 281)

However, perhaps because such components of teaching call for relatively high levels of teacher knowledge and proficiency, they did not feature in the core pedagogical model promulgated to schools.

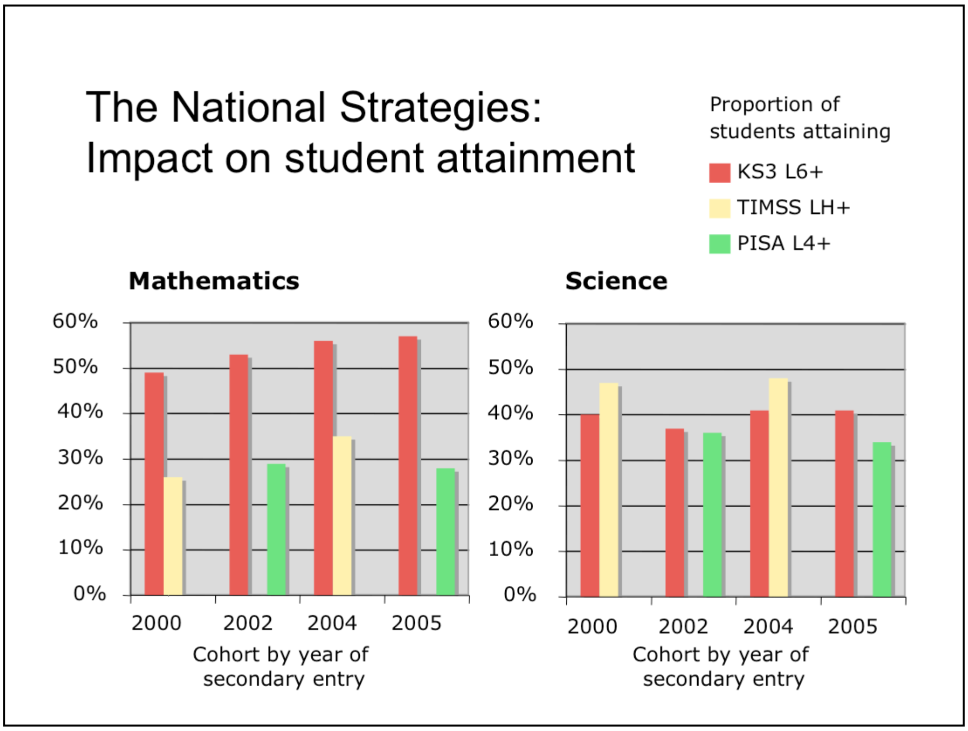
The National Strategies: Key features of the pedagogical model

- Derived from "active teaching" linked to "target setting", placing emphasis on:
  - A detailed schedule of objectives to guide lessons
  - A three-part template for lesson structure
  - Whole-class interaction for pace and progress
  - A system of attainment levels to describe progress
  - Regular target setting, assessment and feedback

The key features of the Strategy's pedagogical model, then, linked tightly structured "interactive teaching" with the "target setting" prominent in outcomes-driven approaches to school improvement. The model emphasised:

- A detailed schedule of objectives to guide lessons;
- A three-part template for lesson structure;
- Whole-class interaction for pace and progress;
- A system of attainment levels to describe progress;
- Regular target setting, assessment and feedback.

(Department for Education and Employment [DfEE], 1998)

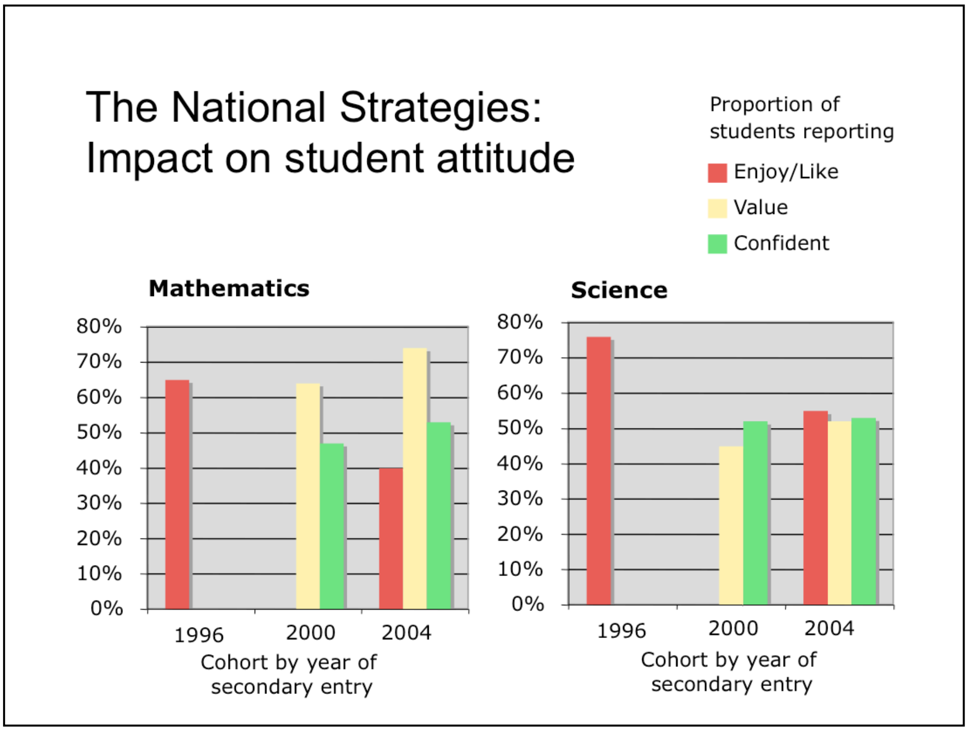


To assess the impact of the Strategies on student attainment we have evidence from three sources: national testing through KS3 SATS (in red); TIMSS tests (in yellow); and PISA tests (in green). Of these, of course, only SATS were administered to every cohort, so the results for TIMSS and PISA are sparser in the graphs. Our focus is on the lower-secondary school: KS3 SATS and TIMSS late in Year 9; and PISA at age 15, typically early in Year 11.

The graphs show the proportion of young people achieving a particular benchmark in each assessment series. The chosen benchmarks are Level 6 in KS3 SATS; the TIMSS "high" benchmark; and PISA Level 4. We need also to remember the changing character of school experience for the successive cohorts: the 2000 cohort was barely affected by the Strategies; the 2002 cohort amongst the first to be touched by them; and the 2004 and 2005 cohorts immersed throughout their primary and lower-secondary schooling.

In Mathematics, outcomes improved considerably by 8 to 9 percentage points over this period both in national testing and TIMSS. However, there was no (statistically significant) change in PISA outcomes. Arguably the difference in results between PISA and TIMSS reflects important differences in the forms of attainment that they measure: in particular, PISA has a distinctive emphasis on mathematical literacy and more functional use of mathematics, not just content knowledge and skills. Indeed, this pattern of partial effectiveness in mathematics of the Strategy pedagogical model matches the strengths and weaknesses acknowledged by its initiators.

In Science all three indicators agree that there was no (statistically significant) change in attainment over this period; this does raise serious questions about the effectiveness in science of the Strategy pedagogical model.



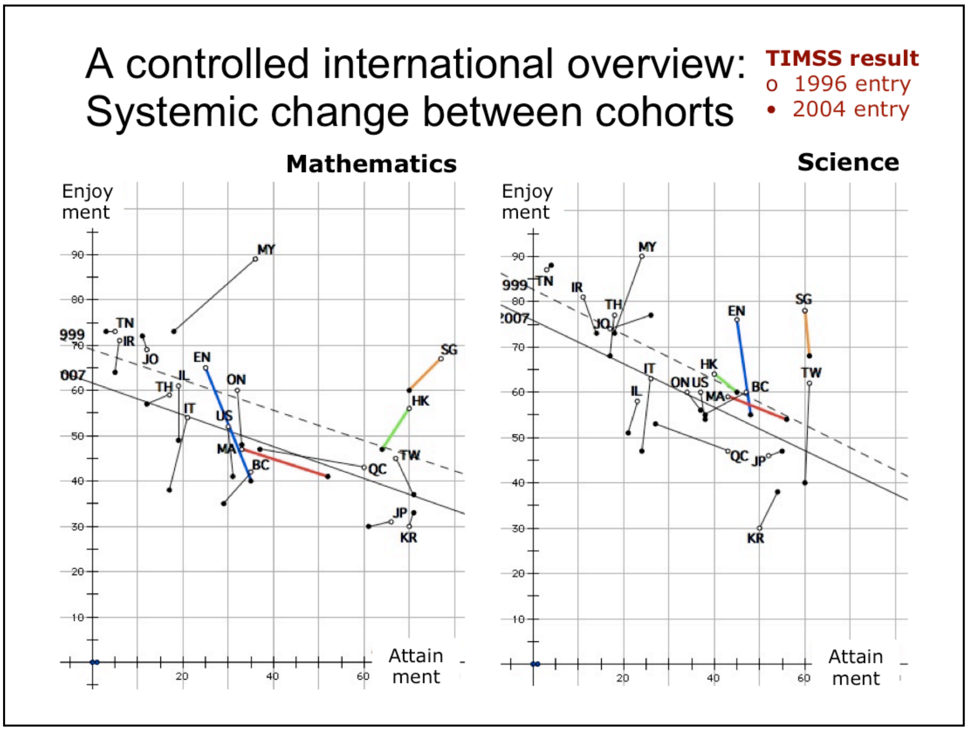
Turning now to student attitude, the only available data series are from TIMSS. The graphs show the proportion of young people achieving the survey benchmark for each of the three facets of attitude surveyed.

Unfortunately the measure concerned with liking the subject and enjoying learning it (in red) was not administered to the 2000 cohort; but there was a very substantial fall in both subjects between the 1996 and and 2004 entry cohorts, of around 25 percentage points in each.

However, the measure concerned with accepting the value of studying the subject (in yellow) shows a rise in both subjects between pre- and post-Strategy cohorts, of 7 to 10 percentage points.

Finally, personal confidence in doing and learning the subject (in green) rises modestly in mathematics, by 6 percentage points, but does not change (statistically significantly) in science.

These trends must raise concerns about impact of the Strategy model on student attitudes: the modest increase in the proportions of students accepting the value of studying the subjects hardly compensates for the marked decline in the proportions liking the subjects and enjoying studying them.



To take a broader international perspective on these trends, these graphs show system-by-system shifts in attainment and enjoyment between the cohorts entering secondary education in 1996 and in 2004. The graphs plot the system averages, for 18 comparable systems including England (in blue) which have consistently participated in TIMSS and have a similar curriculum structure.

There was no overall change in the attainment average across systems: England's 10 percentage point gain in mathematics is well above the norm, and its unchanged performance in science very typical. There was an overall decline in the enjoyment average across systems of 6 to 7 percentage points: England's falls, at around 25 percentage points, are amongst the largest in both the subjects.

In mathematics, the picture for those still high-achieving comparators, Hong-Kong (in green) and Singapore (in ochre), is less encouraging: they have moved back both in student attainment and enjoyment.

The big success is Massachusetts (in red) which has forged forward in attainment in both subjects, while containing falls in enjoyment to below average levels.

A controlled international overview: Systemic reform in Massachusetts

- Systemic improvement programme
  - based on establishing common professional standards and ambitious achievement targets
  - backed by extensive professional development and strong accountability mechanisms
- Normative pedagogical model
  - influenced by more recent research addressing development of higher-order thinking
  - as reflected in *Standards* developed by National Council of Teachers of Mathematics (1989, 2000) and National Academy of Sciences (1995)

Like England, Massachusetts has undertaken a systemic improvement programme since the late 1990s (Driscoll, 2009; Massachusetts Department of Education, 1999). In both systems this has been based on establishing common professional standards and ambitious achievement targets, backed by extensive professional development and strong accountability mechanisms.

Where the two systems differ more is in their normative pedagogical models: In Massachusetts the model has been influenced by more recent research addressing development of higher-order thinking (Massachusetts Department of Education, 1999; Riordan & Noyce, 2001). In particular, the types of approach endorsed by this research have been translated, first, into the *Standards* developed in the US by the National Council of Teachers of Mathematics (1989, 2000) and National Academy of Sciences (1995), and, then, into innovative professional materials and teaching interventions developed with extensive support from the National Science Foundation. Massachusetts has been a national leader in taking up these developments.

Approaches to research synthesis

- Basic pedagogical model for National Strategies formulated through politicised process that filtered out more innovative and recently researched approaches
- Future policy and practice would benefit from taking account of more rigorous approaches to synthesis:
  - Systematic review
  - Best evidence synthesis iteration
  - Meta-analytic approaches

To summarise, then, the basic teaching model promulgated by the National Strategies was formulated through a politicised process that marginalised some facets of teaching known to be pedagogically crucial but professionally challenging. The contrasting TIMSS and PISA trends in mathematics attainment confirm that the Strategy pedagogical model in which "active teaching" was combined with "target setting" was relatively effective in securing mathematical content knowledge and skill, but less so in developing mathematical literacy and functional use of mathematics. The trends in science attainment raise questions about the adequacy of the Strategy model to produce improvement in that subject.

There were contrasting trends in relation to different aspects of attitude to the subjects, where modest increases in the proportions of students accepting the value of studying the subjects was accompanied by marked decline in the proportions and enjoying studying them.

Development of future policy and practice would clearly benefit from rigorous synthesis of available research, and I now turn to that issue. In this second section of the paper I will start by sketching three current approaches to such synthesis: Systematic review; An iterative form of best evidence synthesis; Meta-analysis and a meta-analytic form of best-evidence synthesis.

Approaches to research synthesis: Systematic review

- Programme established by UK Department for Education via the Evidence for Policy and Practice Information and Coordination Centre (EPPI-Centre)
  - Follows standard stages
  - Aims to use explicit, transparent methods
  - Involves range of users to ensure relevance
- Example of the review of *Strategies to raise pupils' motivational effort in Key Stage 4 Mathematics*
  - Very specific focus over limited time period
  - 25 relevant studies identified
  - Only one study provided high weight of evidence

In the UK, a programme of systematic review has been established by the Department for Education through the Evidence for Policy and Practice Information (EPPI) Centre (Bennett et al., 2005; Mathematics Education Review Group, 2009; Science Education Review Group, 2009).

A review of this type follows standard stages which employ explicit transparent methods for the identification, selection, keywording and summarisation of studies. An advisory group involving a range of potential users is involved at every stage of the review to ensure its relevance.

The example of the Mathematics Education Group review of Strategies to raise pupils' motivational effort in Key Stage 4 Mathematics (Kyriacou & Goulding, 2006) is instructive. This review focused very specifically on pupils of mid-below-average to average range of mathematical attainment in England.

It drew on the 25 relevant studies reported in England in the period from September 1999 (when the National Strategies were introduced at primary level) to May 2005. Unfortunately it found only one study that provided a high weight of evidence, meaning that findings had to be tentative because they depended on more extended inference from other studies.

Nevertheless, the review was able to identify key ideas that provided important pointers to potential improvement.

Approaches to research synthesis: Iterative best-evidence synthesis

- Programme established by NZ Ministry of Education
  - Uses research literature to identify what is effective in education for diverse learners
  - Adopts health-of-the-system view that requires dialogue across professional constituencies
- Example of *Effective Pedagogy in Mathematics*
  - Drew on NZ literature complemented by work from other countries with similar characteristics.
  - Identified seminal "landmark" studies to pinpoint how quality teaching might be characterised
  - Derived a common pedagogical principles that appear to hold good across people and settings

Iterative best-evidence synthesis is a programme established by the New Zealand Ministry of Education (Alton-Lee, 2004). It aims to deepen understanding from the research literature of what is effective in education for diverse learners. It adopts a health-of-the-system perspective that requires dialogue with a wide range of professional constituencies.

The example of *Effective Pedagogy in Mathematics/Pàngarau* (Anthony & Walshaw, 2007) drew on NZ literature complemented by reputable work undertaken in other countries with similar characteristics. It identified seminal "landmark" studies that pinpointed in unique ways how quality teaching might be characterised. From these studies it derived common pedagogical principles that appear to hold good across people and settings.

Approaches to research synthesis: Meta-analysis incl. best-evidence synthesis

- Meta-analysis is a well-established approach to summarising studies of the effects of teaching processes on student learning
  - systematically searches for relevant studies and screens them according to explicit criteria
  - classifies the types of teaching process and learning outcome in each accepted study
  - estimates effects through statistical aggregation
- Meta-analytic best-evidence synthesis (Slavin, 1986) adds summary description of each contributing study

Meta-analysis is a well-established approach which summarises studies of the effects of particular types of teaching process on specific aspects of student learning. It systematically searches for relevant studies and screens them according to explicit criteria. It then classifies the types of teaching process investigated and learning outcome measured in each accepted study. Finally, it estimates effects, according to type of teaching approach and learning outcome, through statistical aggregation and modelling.

A variant form of meta-analytic best-evidence synthesis (Slavin, 1986) adds a summary description of each contributing study.

Recent research synthesis on pedagogy: Triangulating the meta-analytic studies

	Schroeder et al. (2007)	*Seidel & Shavelson (2007)*	*Slavin et al. (2008, 2009)*
Subject	Science	Both Ma & Sc	Mathematics
Conceptual framework	Science teaching	Cognitive modelling	Instructional interventions
Teaching construct	Teaching strategies	Learning components	Instructional programs
Field location	Restricted: Only US	Unrestricted: Mainly US, Eur	Unrestricted: Mainly US
Public'n. date	1980-2004	1995-2004	1970-2008
Duration	Unrestricted	Unrestricted	At least 12 wks
Outcomes examined	Achievement	Achievement Attitude	Achievement

Three recent meta-analyses have examined the effective teaching of school mathematics and science.

Schroeder et al. (2007) sought to identify effective teaching strategies in science from studies conducted in the United States.

From a broader meta-analysis of research on effective teaching and learning components, Seidel & Shavelson (2007) reported, as a by-product, findings specifically related to science and mathematics.

Through best-evidence syntheses (of the meta-analytic type) Slavin et al. (2008, 2009) surveyed the effectiveness of specific mathematics programs at the elementary, middle and high school levels.

Although all three meta-analyses examined student attainment, only Seidel & Shavelson considered student attitude.

While it might seem that there is little scope for variation in the conduct of meta-analysis, inspection of these three examples highlights a range of crucial variations in decisions. As the table shows, the conceptual frameworks and teaching constructs employed in each synthesis were rather different. Quite different decisions were also made about the provenance, period and duration of the studies to be considered for inclusion.

Recent research synthesis on pedagogy: Triangulating meta-analytic screening

	Schroeder et al. (2007)	*Seidel & Shavelson (2007)*	*Slavin et al. (2008, 2009)*
Research design(s) accepted	Experimental comparison or evaluation Prior control not required	Correlational survey or experimental comparison Prior control	Randomised or matched experimental comparison Prior control No large gaps
Effect sizes included	Relative Absolute	Relative only	Relative only
Outcome measures accepted	Unrestricted: Generally researcher developed	Unrestricted: Standardised & researcher developed	Screened for intervention bias: Mainly standardised

The reviews also differed in the methodological criteria set for including studies. On research design, only Seidel & Shavelson included correlational surveys as well as experimental studies. In screening experimental studies, Schroeder et al. were most inclusive, and Slavin et al. most exclusive. For example, only Schroeder et al. accepted experimental evaluations (without any control group), and the resulting absolute (rather than relative) effect sizes, whereas Slavin et al. required randomised or matched experimental comparisons. Equally, Slavin et al. barred studies where prior inter-group differences were large, and required appropriate adjustment in accepted studies.

Another important difference is that, whereas the other syntheses accepted a wide range of attainment measures, Slavin et al. rejected comparisons based on aspects likely to have received little or no attention in control groups. Hence the studies included by Slavin et al. predominantly employed standardised tests and state assessments. While this ensured that there was no form of bias towards the intervention, it may also have introduced another type of bias, towards measures emphasising relatively basic content knowledge and skills.

There is a surprising lack of overlap in studies included in the three reviews: the most striking illustration is that none of the 32 studies included by Schroeder et al. which were eligible, by publication date, for the Seidel & Shavelson synthesis, featured in the latter. Finally, there proved to be substantial divergences in the classification of those studies included in both the Slavin et al. and Seidel & Shavelson syntheses.

Recent research synthesis on pedagogy: Meta-analytic findings on attainment effects

	Mathematics		**Science**	
	Slavin et al.	*Seidel & Shavelson*	*Seidel & Shavelson*	*Schroeder et al.*
Domain-specific inquiry	No cognate category	0.37 [22]	0.63 [28]	0.65 [12]
Co-operative groupwork	0.36 [17]	-0.04 [42]	0.00 [35]	0.96 [3]
Enhanced context	No cognate category	No cognate category	No cognate category	1.48 [6]
Active teaching	0.43 [10]	No cognate category	No cognate category	—— [0]

Four pedagogical constructs emerged as particularly promising. However, the differing conceptual frameworks employed by the the meta-analyses mean that not all contain categories cognate to every one of these constructs.

The cells for which effects are reported show an average effect size followed [in brackets] by the number of studies on which it is based. [Technical note: The effect sizes quoted by Seidel & Shavelson have been transformed from Fisher's Z metric to Cohen's d to make them directly comparable with the other syntheses.]

For attainment, the two reviews which featured Domain-specific inquiry (namely extended forms of mathematical problem solving or scientific inquiry) agreed on the strong effectiveness of this type across subjects.

However, while all three reviews contained a category of Cooperative groupwork, they disagreed on its effectiveness: one potentially crucial factor appears to be whether participating teachers had undergone relevant professional development (which Slavin et al. required for a study to be included in this category).

Only Schroeder et al contained a clear-cut category of Enhanced context (namely teaching which relates learning to students' experiences or interests). It found this to be highly effective (in science).

Schroeder et al. looked for studies of Direct instruction in science but found none. Slavin et al. contained two relevant categories: Direct instruction and Classroom management and motivation (which included influential studies of active teaching). These have been combined under the head of Active teaching since both categories were found effective at comparable levels.

Recent research synthesis on pedagogy: Meta-analytic findings on attitude effects

	Mathematics	Science
	Seidel & Shavelson	*Seidel & Shavelson*
Domain-specific inquiry	—— [<5]	0.35 [7]
Co-operative groupwork	0.02 [9]	0.41 [14]
Enhanced context	No cognate category	No cognate category
Active teaching	No cognate category	No cognate category

Only Seidel & Shavelson examined attitude effects. The evidence base was much smaller than for attainment effects; in particular, insufficient studies were found to estimate an effect in one key cell.

The picture is more encouraging in science where the components of Domain-specific inquiry and Cooperative groupwork both emerge as effective. The picture in mathematics is disappointing since no effective component emerges.

Recent research synthesis: Triangulating wider findings on identity and attitude

BES iteration on effective mathematics teaching

- Teaching that takes students' mathematical thinking seriously is underpinned by principles such as:
  - all students being empowered to develop mathematical identities and knowledge
  - interpersonal respect, sensitivity, fairness and consistency being shown to all students

EPPI review on motivational effort in mathematics

- Effective pedagogical strategies are characterised by:
  - a caring and supportive classroom climate
  - activities which pupils find challenging and enjoyable
  - pupils gaining deeper understanding of mathematics
  - opportunities for pupils to collaborate

Now I will triangulate findings from the meta-analytic review against the BES and EPPI syntheses, focusing on two major areas.

In mathematics in particular, the meta-analyses were not able to assess the effectiveness of Domain-specific inquiry on attitudes towards learning mathematics, and were discouraging about the effectiveness of Co-operative groupwork on such attitudes. Both the BES iteration on effective mathematics teaching and the EPPI review on raising motivational effort in mathematics link positive attitude to the the development of constructive identity in relation to the subject. Both also point to the importance of establishing high and equitable expectations of students' potential and a positive psychosocial climate around classroom mathematical activity. The EPPI review also concludes that providing opportunities for pupils to collaborate can raise students' motivational effort in mathematics.

It seems, then, that there is potential for the pedagogical methods (and associated mechanisms) that have proved effective in generating positive attitudes towards science learning to be effective in mathematics, as long as expectations of students' potential are suitably ambitious and classroom ethos supportive to achieving them.

Recent research synthesis: Triangulating wider findings on co-operative groupwork

BES iteration on effective mathematics teaching

- Small-group work can support engagement
- Students may need opportunities to think quietly
- Many students are reluctant to share their thinking

EPPI review of group discussions in science teaching

- Students often struggle to express coherent arguments, and demonstrate a low level of engagement with tasks
- Groups function best, understanding improves most:
  – with groups constituted so that differing views voiced
  – when students receive training on group processes
  – when "cues" support the structuring of discussions

As regards the value of cooperative groupwork in mathematics, the further syntheses confirm a lack of conclusive results.

Whereas the EPPI review on raising motivational effort in mathematics identified collaborative activity as a means of engendering motivational effort in mathematics, the BES iteration on effective mathematics teaching was more reserved. It reports that while some researchers have found that small-group work can provide the context for social and cognitive engagement, others have cautioned that students need opportunities and time to think and work quietly. It also noted that many students, including limited-English-speaking students, are reluctant to share their thinking in class discussions. The EPPI review of small group discussions in science teaching also notes some reservations, reporting that, during small group discussion, students often struggle to formulate and express coherent arguments, and demonstrate a low level of engagement with tasks. However, this EPPI review reports that preparation and organisation are crucial to the success of small group work. Groups function more purposefully, and understanding improves most:

- when groups are constituted so that differing views are voiced;
- when students receive training on effective group work;
- when "cues" support the structuring of discussions

(Bennett et al., 2010).

Research-informed conclusions for pedagogical improvement at scale

- In mathematics, varied sources agree that the active teaching model promoted by the Strategies is effective in securing content knowledge and skills but less so in developing higher-order and functional thinking
- There is little support for this teaching model in science
- Domain-specific enquiry that takes students' thinking seriously strengthens attainment and (plausibly) attitude
- Co-operative groupwork strengthens attainment and (at least in science) attitude, as long as students are properly prepared and activity well supported
- Enhanced context, linked to student experiences and interests, is beneficial (at least for science attainment)

We have seen that established practice in England is built around the objectives-driven interactive teaching model promoted by the Strategies. In mathematics, both the research literature and trends in English performance in international comparisons suggest that this model is reasonably effective in securing basic knowledge but less so in developing more flexible higher-order thinking.

In science, there appears to be little evidence in support of this model.

The research literature (and the related pedagogical developments underpinning the Massachusetts success identified from international series) indicate that use of:

- domain-specific enquiry that takes students' thinking seriously would strengthen attainment and (provenly in science, plausibly in mathematics) attitude;
- co-operative groupwork would strengthen achievement and (at least in science) attitude, as long as students and teachers were properly prepared and activity well structured;
- enhanced context, linked to the interests and experiences of students, could be beneficial (provenly at least for science attainment).

References

Alton-Lee, A. (2004). *Guidelines For Generating a Best Evidence Synthesis Iteration.* Wellington NZ: Ministry of Education.

Anthony, G. & Walshaw, M. (2007). *Effective Pedagogy in Mathematics/Pangarau: Best Evidence Synthesis Iteration.* Wellington NZ: Ministry of Education.

Bennett, J., Hogarth, S., Lubben, F., Campbell, B. & Robinson, A. (2010). Talking science: The research evidence on the use of small group discussions in science teaching. *International Journal of Science Education,* 32(1), 69-95.

Bennett, J., Lubben, F., Hogarth, S. & Campbell, B. (2005). Systematic reviews of research in science education: rigour or rigidity? *International Journal of Science Education,* 27(4), 387-406.

Bradshaw, J., Ager, R., Burge, B., & Wheater, R. (2010). PISA 2009: Achievement of 15-year-olds in England. Slough: NFER.

Department for Children, Schools and Families [DCSF] (2009). National Curriculum Assessments at Key Stage 3 in England, 2008. *Statistical First Release* 10/2009. London: DCSF.

Department for Education and Employment [DfEE] (1998). *The Implementation of the National Numeracy Strategy: The Final Report of the Numeracy Task Force.* London: DfEE.

Driscoll, D. (2009). Mathematics reform: Lessons learned in Massachusetts. Presentation at Indiana Mathematics Summit. Accessed 28 August 2009, http://media.doe.in.gov/math/2009Summit.html

Good, T.L., Grouws, D.A. & Ebmeier, H. (1983). *Active mathematics teaching.* New York: Longman.

Hipkins, R. Bolstad, R., Baker, R., Jones, A., Barker, M. , Bell, B., Coll, R., Cooper, B., Forret, M., Harlow, A., Taylor, I., France, B. & Haigh, M. (2002). *Curriculum, Learning and Effective Pedagogy: A Literature Review in Science Education.* Wellington NZ: Ministry of Education.

Kyriacou, C. & Goulding, M. (2006). *A systematic review of strategies to raise pupils' motivational effort in Key Stage 4 Mathematics.* London: EPPI-Centre.

Massachusetts Department of Education (1999). *Partnerships Advancing the Learning of Mathematics and Science.* Malden MA: Mass. Dept. of Ed.

Mathematics Education Review Group, 2009. Published reviews. Accessed 15 December 2009, http://eppi.ioe.ac.uk/cms/Default.aspx?tabid=405

Mullis, I., Martin, M. & Foy, P. (2008a). *TIMSS 2007 International Mathematics Report: Findings from IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades.* Boston: TIMSS International Study Center.

Mullis, I., Martin, M. & Foy, P. (2008b). *TIMSS 2007 International Science Report: Findings from IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades.* Boston: TIMSS International Study Center.

National Academy of Sciences (1995). *National Science Education Standards.* Washington DC: National Academies Press.

National Council of Teachers of Mathematics (1989). *Standards for School Mathematics.* Reston VA: NCTM.

National Council of Teachers of Mathematics (2000). *Principles and Standards for School Math.* Reston VA: NCTM.

Reynolds, D. & Muijs, R.D. (1999). The effective teaching of mathematics: a review of research. *School Leadership and Management* 19(3), 273-88.

Riordan, J. & Noyce, P. (2001). The impact of two *Standards*-based mathematics curricula on student achievement in Massachusetts. *Journal for Research in Mathematics Education,* 32(4), 368-398.

Ruthven, K. (2011). Using international study series and meta-analytic research syntheses to scope pedagogical development aimed at improving student attitude and achievement in school mathematics and science. *International Journal of Science and Mathematics Education*, 9(2), 419-458.

Schroeder, C. M., Scott, T. P., Tolson, H., Huang, T.-Y. & Lee, Y.-H. (2007). A meta-analysis of national research: Effects of teaching strategies on student achievement in science in the United States. *Journal of Research in Science Teaching,* 44(10), 1436-1460.

Science Education Review Group (2009). Published reviews. Accessed 15 December 2009, http://eppi.ioe.ac.uk/cms/Default.aspx?tabid=444

Seidel, T. & Shavelson, R. J. (2007). Teaching effectiveness research in the past decade: The role of theory and research design in disentangling meta-analysis research. *Review of Educational Research,* 77(4), 454-499.

Slavin, R. (1986). Best-evidence synthesis: An alternative to meta-analytic and traditional reviews. *Educational Researcher,* 15(9), 5-11.

Slavin, R. (2008). What works? Issues in synthesizing educational program evaluations. *Educational Researcher,* 37(1), 3-14.

Slavin, R. & Lake, C. (2008). Effective programs in elementary mathematics. *Review of Educational Research,* 78(3), 427-515.

Slavin, R., Lake, C. & Groff, C. (2009). Effective programs in middle and high school mathematics. *Review of Educational Research,* 79(2), 839-911.

Sturman, L., Ruddock, G., Burge, B., Styles, B., Lin, Y. & Vappula, H. (2008). *England's Achievement in TIMSS 2007: National Report for England*. Slough: NFER.