

KENNETH RUTHVEN

USING INTERNATIONAL STUDY SERIES AND META-ANALYTIC RESEARCH SYNTHESSES TO SCOPE PEDAGOGICAL DEVELOPMENT AIMED AT IMPROVING STUDENT ATTITUDE AND ACHIEVEMENT IN SCHOOL MATHEMATICS AND SCIENCE

ABSTRACT: Taking lower-secondary schooling within the English educational system as an example, this paper illustrates the contribution of two bodies of international scholarship to the scoping of research-based pedagogical development that aims to improve student attitude and achievement in science and mathematics. After sketching the English context of systemic reform, the paper uses findings from the TIMSS international study series to illuminate performance trends, by analysing them within a framework of cross-system and between-subject comparison. Contrary to the optimistic picture from national assessment, the TIMSS findings suggest that systemic reform in England has produced fundamental gains only in student achievement in mathematics, and serious declines in student attitude towards both mathematics and science. Prompted by more favourable patterns elsewhere, the paper then triangulates the findings of recent meta-analytic research syntheses to identify promising lines of pedagogical development. Despite important differences in the conceptual frameworks and analytic methods of these syntheses, reasonably robust conclusions can be drawn about the effectiveness of four teaching components: *domain-specific inquiry* for student achievement in both subjects, student attitude in science, and learning processes in mathematics; *cooperative groupwork* for learning and attitude in science; *contextual orientation* for achievement in science; and *active teaching* for achievement in mathematics. Equally, discrepancies between findings or insufficiencies of evidence highlight a number of impacts particularly deserving deeper analysis or further investigation: *cooperative groupwork* on achievement outcomes; differing forms of *learning assessment* on both attitude and achievement outcomes; *contextual orientation* on outcomes in mathematics; and *active teaching* on outcomes in science.

KEYWORDS: effective teaching; instructional improvement; international comparisons; research reviews; teaching methods; mathematics teaching; science teaching; England; TIMSS.

INTRODUCTION

The idea of research-informed and evidence-based improvement of education is becoming increasingly influential around the world. A current British initiative has been sponsored by the UK Economic and Social Research Council (ESRC, 2006) and other national agencies with interests in STEM education. It seeks to capitalise on previous research and undertake further studies with the intention of guiding changes in educational policy and practice to enhance young people's school achievement in science and mathematics, and to significantly increase their participation in further study and employment in these areas. Within this initiative, the *Effecting Principled Improvement in STEM Education (epiSTEMe)* project (Ruthven et al., 2010) is concerned with research-based pedagogical development aimed at improving student engagement and achievement in early secondary-school physical science and mathematics, in ways suited to implementation at scale within the English educational system¹.

This paper will illustrate the contribution of two bodies of international research to the scoping of pedagogical development of this type. Although the paper has a specific focus on the English educational system and the lower-secondary phase, it has been written in a way which seeks to treat these simply as convenient sites for examining issues of wider interest and significance. The paper begins by sketching the main features of recent systemic reform aimed at improving mathematics and science teaching in English schools, and the influence of earlier pedagogical research on this reform. The paper then shows how findings from the TIMSS international study series can be used to examine trends in the performance of an educational system within a framework of cross-system and between-subject comparison, providing a form of independent evaluation of the English reform. The picture that emerges is a mixed one. Although the reform has succeeded in improving some aspects of student achievement, this has been at the expense of student attitude. More fundamentally, while reform has successfully implanted a pedagogical model influenced by earlier research on effective skills-focused teaching, only limited account has been taken of more recent research addressing a broader educational agenda. In particular, the English system has been outperformed by another that has followed a programme of systemic reform more cognisant of this recent research. The paper proceeds, then, to examine the findings of recent research syntheses on mathematics and science teaching, focusing particularly on meta-analyses of teaching effects, with a view to identifying promising lines of pedagogical development.

AN ENGLISH SYSTEMIC INITIATIVE TO ESTABLISH A RESEARCH-BASED MODEL FOR EFFECTIVE TEACHING

Since the late 1980s, schooling in England has been reshaped by a series of government initiatives aimed at standardising educational provision and improving educational outcomes. From 1989, a national curriculum was introduced, in tandem with a system of national assessment covering the primary and lower-secondary phases, accompanied by a strengthening of mechanisms of professional accountability, notably through teacher evaluation and school inspection. From 1997, these policies were extended by setting targets for improved school and teacher performance, particularly in terms of student achievement, and by launching a national programme of school improvement, supported by extensive professional development, which established official pedagogical norms for the primary and lower-secondary phases. This programme gave special attention to the core competences of literacy (within the subject of English) and numeracy (within the subject of mathematics). From 1999, in particular, the teaching of mathematics in primary schools in England was expected to follow the pedagogical model established by the *National Numeracy Strategy*.

Educational research had a strong influence on the development of this pedagogical model (Brown, Askew, Millett & Rhodes, 2003), even if significant differences emerged between “school improvers” and “mathematics educators” (Ruthven, 2008). The chair of the expert Task Force which formulated the model described it as being based on “an agreed ‘solid centre’ of practice” (Reynolds & Muijs, 1999, p. 273) – although this was contested (Brown et al., 1998). The principal research source was an American tradition of “process-product” research on effective mathematics teaching, the main findings of which were claimed to accord both with the much smaller body of relevant British research, and with the judgement of English school inspectors in their contemporary reports on the school system. It was work on the construct of “direct instruction” or “active teaching” within this American tradition – notably by the Missouri Mathematics Program (Good, Grouws, & Ebmeier, 1983) –

that provided the main research backing for “the whole-class ‘interactive’ model of maths teaching” (Reynolds & Muijs, 1999, p. 274) promoted by the *Strategy*. Conceding, however, that this model had been validated primarily in relation to the teaching of basic skills, the relevance of more recent research on the development of higher-order thinking was acknowledged:

[A] number of additional classroom processes may be needed to enhance higher order thinking: a focus on meaning and understanding in mathematics, direct teaching of higher level cognitive strategies and problem-solving, and co-operative small group work. (Reynolds & Muijs, 1999, p. 281)

The key features of the model prescribed by the *Strategy* related to lesson planning, curriculum coverage, teaching method, and classroom assessment. First, the *Strategy* provided a detailed schedule of teaching objectives and examples for each year group, intended to guide day-to-day lesson planning and ensure coverage of the national curriculum.

Second, the *Strategy* promoted use of a three-part template for lesson organisation:

- 5 to 10 minutes whole-class “mental and oral work to rehearse and sharpen skills”.
- 30 to 40 minutes of whole-class then group/pair/individual work based on:
 - “clear objectives shared with pupils”;
 - “interactive/direct teaching input”;
 - “practical and written work for all the class”;
 - “continued interaction and intervention [by the teacher]”.
- 10 to 15 minutes of whole-class plenary based on:
 - “feedback from children to identify progress and sort misconceptions”;
 - “summary of key ideas, what to remember”;
 - “links made to other work, next steps discussed”.

(DfEE, 1998, p. 18)

Third, the *Strategy* emphasised the development of explicit assessment processes, particularly with a view to:

- “agreeing personal targets with each pupil, and discussing and reviewing their progress towards them”;
- “giving constructive feedback which will enable pupils to improve their strategies”.

(DfEE, 1998, p. 59)

Initially developed for use at primary level, this pedagogical approach was extended to lower-secondary level for the teaching of mathematics (from the student cohort entering in 2001) and science (from 2002). Reporting on implementation at lower-secondary level in mathematics, the main trends noted by school inspectors were towards “improvements in the planning of teaching, with a greater focus on learning objectives, the structure of lessons and teachers’ use of questioning”, but “insufficient emphasis on using independent, collaborative and oral work to encourage pupils to grapple with ideas” (OfStEd, 2004, pp. 21 & 23). The inspectors also noted the introduction of “systems... for regular monitoring of pupils’ performance, with action taken to help them improve”, extending in some schools to “specific curricular targets [being] routinely set for individual pupils” (OfStEd, 2004, p. 24). The approach developed in science was similar in many respects. At lower-secondary level, inspectors reported that a three-part lesson structure was near universal, as was emphasis on teaching to explicit learning objectives, but with teachers making efforts “to build on pupils’ prior experience, most often through questioning” (OfStEd, 2004, p. 31).

Likewise, inspectors reported that teacher assessment was “being increasingly used more successfully to guide pupils on how to improve their work, although more effective use of target-setting for individual pupils [wa]s needed” (OfStEd, 2004, p.29).

This, then, is the background against which the proposal for the *epiSTEME* project was formulated in early 2007. However, by the time that the project began in late 2008, important new information had become available about the impact of the more recent wave of reform, as will now be discussed.

USING THE TIMSS STUDY SERIES TO ANALYSE CHANGES IN SYSTEM PERFORMANCE

In late 2008, publication of a new round of results from the Trends in International Mathematics and Science Study [TIMSS] provided fresh information on the performance of the English educational system, permitting more independent evaluation of the impact of the second wave of reform within a broader framework that makes possible cross-system and between-subject comparison. An important advantage of such international study series is that their estimates of student achievement are less open to the risk of inflation as a result of the strategic “teaching (directly) to the (predictable) test” that high-stakes national testing encourages. A further advantage is that they provide assessments not just of student achievement but of student attitude, broadening the types of outcome available for consideration.

Compared to the PISA study series, the achievement measures of TIMSS are more curriculum based, and so better indicators of the developing knowledge-base necessary for advanced study (Ruddock et al., 2006); important for the particular concerns of the ESRC initiative with raising student participation in more advanced mathematical and scientific courses².

Equally, the TIMSS series normally allows an age cohort to be tracked from the elementary/primary level (Grade 4 internationally/Year 5 in England) in one study to the middle/lower-secondary level (Grade 8/Year 9) in the next. Thus (as shown in Table 1) the TIMSS 1995 Grade 4 cohort had entered the primary phase in 1990 and moved on to the lower-secondary phase in 1996, becoming (more or less) the TIMSS 1999 Grade 8 cohort (represented, of course, by a different random sample); we will refer to this as the “earlier” cohort. The entire schooling of this earlier cohort took place following the first wave of reform but well prior to the second wave associated with the *Strategy*. TIMSS provides less information about the “intermediate” cohort, the last to have virtually no experience under the second wave of reform³. It is the “later” cohort, entering the primary phase in 1998, moving on to the lower-secondary phase in 2004, and surveyed at Grade 8 by TIMSS in 2007, that was entirely schooled under the national *Strategy*. Thus any improvement in system outcomes as a result of the second wave of reforms should be indicated by changes between the earlier and intermediate cohorts on the one hand, and the later cohort on the other.

TABLE 1: Progress of student cohorts through school phases and the TIMSS surveys

Cohort	Entered primary phase	Surveyed at Grade 4 (Year 5)	Entered secondary phase	Surveyed at Grade 8 (Year 9)
Earlier cohort	1990	1995	1996	1999
Intermediate cohort	1994	<i>not surveyed</i>	2000	2003
Later cohort	1998	2003	2004	2007

Information and data used in this paper have been extracted from the most recent TIMSS reports (Mullis, Martin & Foy 2008a; 2008b; Sturman et al., 2008). Accordingly, references here to the statistical significance of results reflect the analyses reported in these documents. The findings for England will be situated within the distribution of results across those 18 educational systems that participated in both the 1999 and 2007 TIMSS studies, and which, like England, taught a general/integrated science curriculum at this level⁴. While our main concern will be with change in performance at the Grade 8 level, information about Grade 4 may also be of value in interpreting such change.

Establishing Benchmarks for Student Achievement and Attitude

TIMSS defines (in terms of test scores) four “international benchmarks” for student achievement (ranging from “advanced” to “low”), and reports the proportion of students in each system achieving each of these benchmarks. Here, the focus will be on what TIMSS terms the “high” benchmark. In mathematics at Grade 8 level, this benchmark is characterised in terms of students being able to “apply their understanding and knowledge in a variety of relatively complex situations”, compared to the lower “intermediate” benchmark where students “apply basic mathematical knowledge in straightforward situations”. This “high” benchmark provides, then, a good marker for the level of capability required for longer-term progression into more advanced study in STEM fields. Likewise, in science, students who achieve the “high” benchmark at Grade 8 level are characterised as being able to “demonstrate conceptual understanding of some [science]”, compared to the “intermediate” benchmark where students “recognize and communicate basic scientific knowledge”. Thus, in what follows, the operational index of subject achievement will be the percentage of students reaching the high achievement benchmark in that subject.

Turning from achievement in a subject to attitude towards it, the most relevant outcome measure available in the TIMSS study series is the proportion of students displaying “high positive affect towards [subject]”. We can treat this as a benchmark of attitude paralleling the benchmark of achievement analysed above. Operationally, to figure in this category, students had to respond affirmatively (on average) to three statements (agreeing that they like the subject, that they enjoy learning the subject, and disagreeing that the subject is boring). This benchmark provides, then, a suitable marker for the kinds of attitude conducive to longer-term progression into more advanced study in STEM fields. Thus, in what follows, the operational index of subject attitude is the percentage of students reaching this positive attitude benchmark in that subject.

This apparatus will now be used to analyse trends in system performance. Each set of index scores from the 18 systems creates a distribution. For example, in Table 2 below, the proportion of English students reaching the high achievement benchmark for science was 45% in 1999 and 48% in 2007, a (statistically nonsignificant) rise of 3%. Across the 18 systems as a whole, the median proportion was 39% in 1999 and 38% in 2007, and the median change in proportion a rise of one percentage point. While the main analysis will focus on these cohorts because fuller information is available on them, evidence about the intermediate cohort will later help to triangulate and sharpen findings on achievement.

Examining Between-Cohort Changes in System Performance on Student Achievement

Table 2 shows that, between the earlier and later cohorts, there was little overall change in levels of science achievement across systems. In absolute terms, England's performance did not change significantly; and in relative terms, England remained well above the median position and a little below the upper quartile.

TABLE 2: Trends in system performance in Grade 8 science achievement: Proportion of students reaching the TIMSS high achievement benchmark

Distribution	Earlier cohort scores	Later cohort scores	Between cohort changes
Upper quartile	47%	53%	+ 3
Median	39%	38%	+ 1
Lower quartile	23%	22%	- 2
England	45%	48%	+ 3^{NS}

^{NS} later cohort not significantly different

TABLE 3: Trends in system performance in Grade 8 mathematics achievement: proportion of students reaching the TIMSS high achievement benchmark

Distribution	Earlier cohort scores	Later cohort scores	Between cohort changes
Upper quartile	65%	59%	+ 1
Median	33%	32%	- 2
Lower quartile	20%	17%	- 6
England	25%	35%	+ 10[↑]

[↑] later cohort significantly higher

The situation as regards mathematics is rather different. Table 3 suggests that overall levels of mathematics achievement across systems tended to drift slightly downwards between the cohorts. However, there was a marked improvement (the second largest such improvement) in the absolute performance of England. In relative terms, England moved from well below the median position (closer indeed to the lower quartile) to a little above the median.

Comparing the two subjects, then, England performs less strongly in mathematics than in science relative to the other systems. However, whereas England's performance in science changed little between the two cohorts, performance in mathematics improved, although not to a level matching science.

FIGURE 1: Change in Grade 8 (Year 9) science achievement between English 1999, 2003 and 2007 cohorts: Proportion of students achieving the TIMSS high benchmark, compared to proportion achieving level 6 in national KS3 tests.

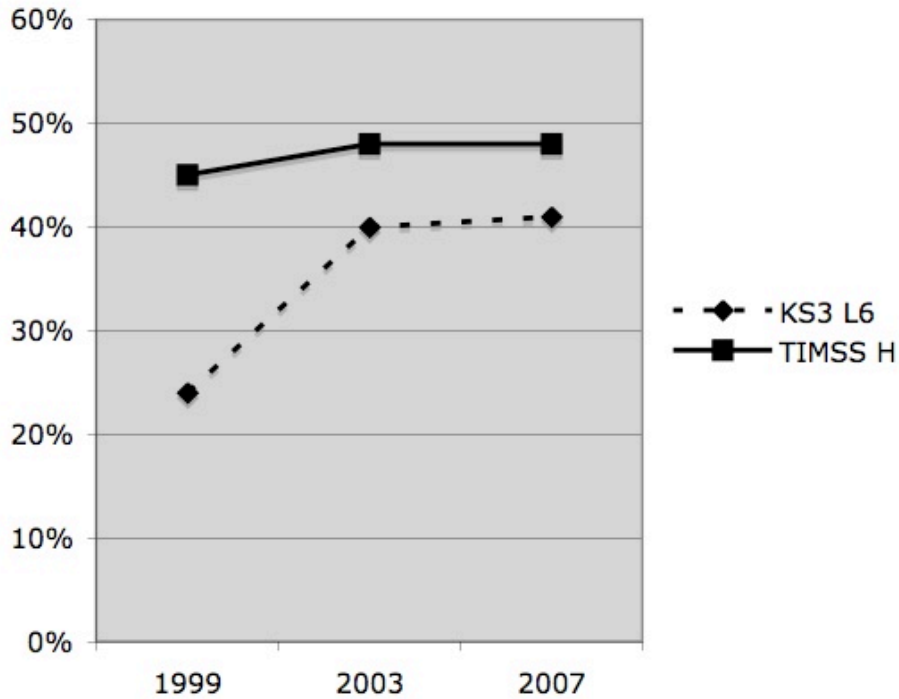
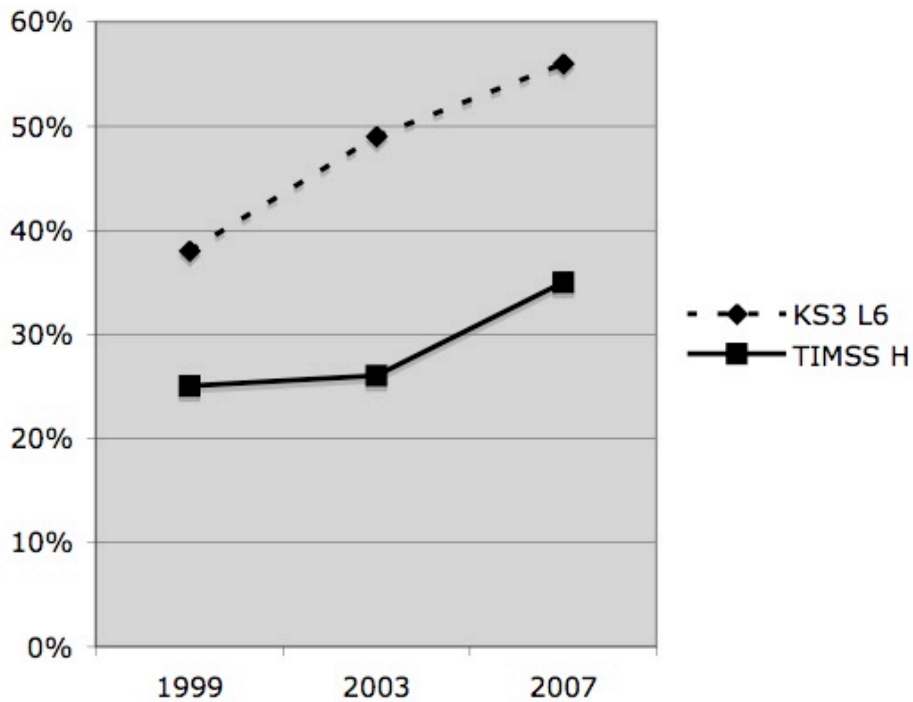


FIGURE 2: Change in Grade 8 (Year 9) mathematics achievement between English 1999, 2003 and 2007 cohorts: Proportion of students achieving the TIMSS high benchmark, compared to proportion achieving level 6 in national KS3 tests.



To what degree, though, might this rise in mathematics performance be a legacy of improvement during the primary years (to Grade 4/Year 5) rather than indicative of enhancement over the middle/lower secondary years (to Grade 8/Year 9)? Unfortunately, TIMSS data is available at Grade 4 level for only 8 of these 18 systems. Of those 8 systems, England had the largest improvement in mathematics achievement at both levels. Nevertheless, the English improvement at Grade 8 level (of 10 percentage points) was notably smaller than the improvement at Grade 4 level (where the rise was 19 percentage points, from 24% to 43%). Given the magnitude of this improvement at Grade 4 level, it is plausible to conjecture that it contributed to the subsequent improvement at Grade 8 level⁵.

Triangulating Between-Cohort Changes in System Performance on Student Achievement

These TIMSS findings make it possible to triangulate evidence from national assessment (DCSF, 2008). Over this period, English schools and teachers were under enormous pressure to improve the performance of students in national tests: for lower-secondary assessment at the end of Year 9, level 6 represented the key benchmark of higher achievement. The graphs for science (Figure 1) and mathematics (Figure 2) show that the relative demands of the two benchmarks differ between subjects: in science, the level 6 benchmark in national testing is more demanding than the TIMSS high achievement benchmark; in mathematics, this pattern is reversed. In TIMSS terms, then, the expectations of English national assessment are higher in science than in mathematics.

Turning to trends over time, the graphs show that between 1999 and 2003, performance on national tests improved markedly in both science (Figure 1) and mathematics (Figure 2), while performance on TIMSS remained static: this points to a relatively narrow form of improvement in which schools and teachers were becoming increasingly effective in enhancing student performance specifically on national tests. Between 2003 and 2007, however, trends for the two subjects diverged. Both in national tests and in TIMSS, performance remained (statistically) static in science, whereas it rose markedly in mathematics. This suggests that the second wave of reform (affecting only the later cohort) led to fundamental improvement in this aspect of performance in mathematics, but not in science⁶.

Examining Between-Cohort Changes in System Performance on Student Attitude

Turning to assessment of student attitude within TIMSS, Table 4 shows a general downward shift in attitude to science between the two cohorts. England displays a marked absolute decline (the second largest of any system): in relative terms, it moved from just below the upper quartile to the median position (which was quite close to the lower quartile).

Likewise, Table 5 shows a general downward shift in attitude to mathematics between the two cohorts. However, England displayed a particularly marked absolute decline in performance (the largest of any system): in relative terms, it moved from just below the upper quartile position to just above the lower quartile.

Comparing the two subjects, then, the English system follows the international trend to perform more strongly on attitude in science than in mathematics. In both subjects, however, England saw a very substantial fall in attitude between the two cohorts, markedly greater than the international trend.

TABLE 4: Trends in system performance in Grade 8 science attitude: proportion of students reaching the TIMSS positive attitude benchmark

Distribution	Earlier cohort scores	Later cohort scores	Between cohort changes
Upper quartile	77%	68%	+ 1
Median	63%	55%	- 6
Lower quartile	59%	52%	- 10
England	76%	55%	- 21 ↓

↓ later cohort significantly lower

TABLE 5: Trends in system performance in Grade 8 mathematics attitude: proportion of students reaching the TIMSS positive attitude benchmark

Distribution	Earlier cohort scores	Later cohort scores	Between cohort changes
Upper quartile	67%	59%	- 1
Median	58%	47%	- 7
Lower quartile	46%	39%	- 12
England	65%	40%	- 25 ↓

↓ later cohort significantly lower

Examining Between-Cohort Changes in System Performance on Combined Outcomes

Using a scatterplot makes it easier to examine combined changes in the performance of each system on achievement and attitude outcomes; in science (Figure 3), and in mathematics (Figure 4). In some systems, there was very little change between cohorts: in Tunisia (TN) for example (at the top left of the two figures), there were negligible changes in achievement or attitude in either subject. In others, Malaysia (MY) for example (again towards the top left of the two figures), there were marked falls in both achievement and attitude in both subjects.

In both subjects, and for both cohorts, trend lines relating attitude to achievement across the systems as a whole indicate that, at this system level, higher performance on achievement tends to be associated with lower performance on attitude; this is likely to reflect underlying system-level differences (in factors such as economic development and educational orientation) that mediate both achievement and attitude. The shift between the trend-lines from 1999 (dashed) to 2007 (solid) reflects some form of decline between the two cohorts; and the segments indicating movement of individual systems show that this is predominantly due to changes (downwards) in attitude rather than (backwards) in achievement.

FIGURE 3: International trends in student achievement and attitude in Grade 8 science: Change in system performances between 1999 and 2007 TIMSS cohorts

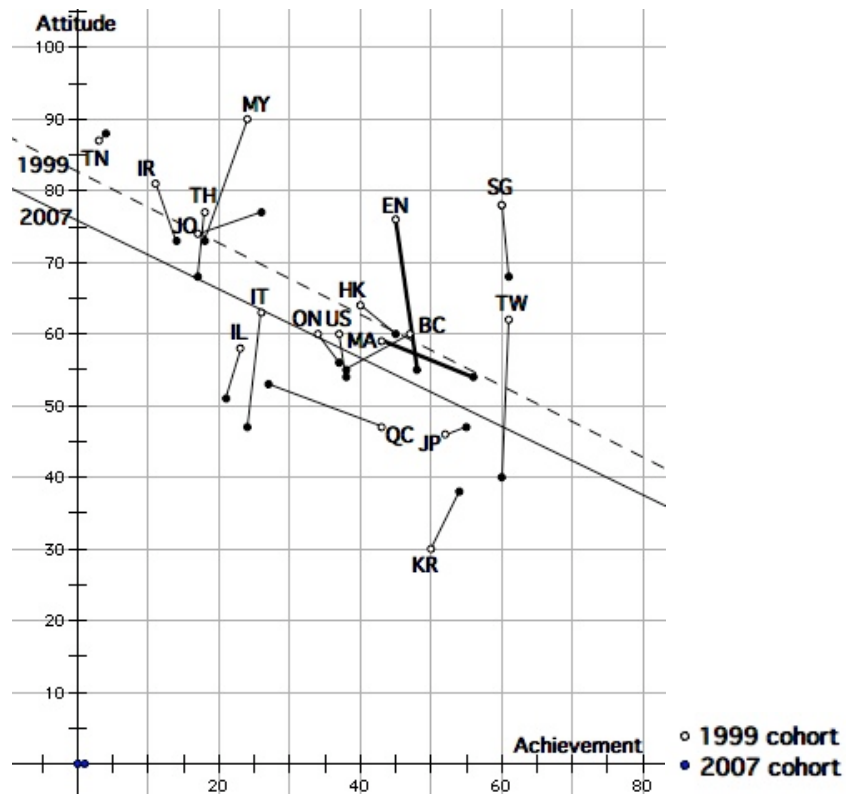
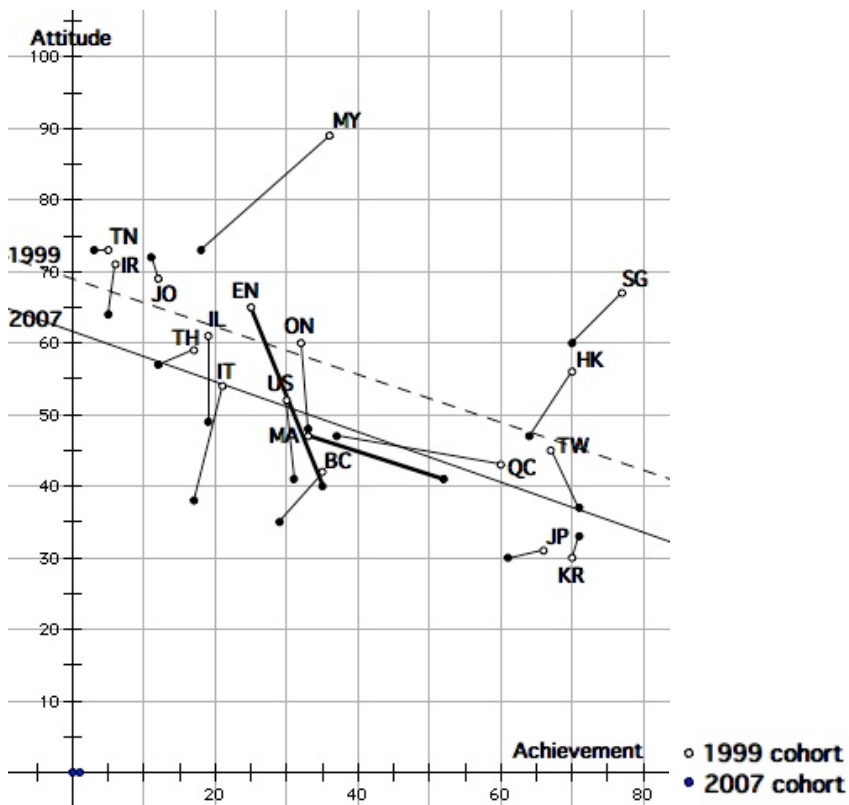


FIGURE 4: International trends in student achievement and attitude in Grade 8 mathematics: Change in system performances between 1999 and 2007 TIMSS cohorts



On achievement, the consistent strong riser is Massachusetts (MA), which improved 13 percentage points in science (from 43% to 56%), and 19 percentage points in mathematics (from 33% to 52%). On attitude, Massachusetts declined 5 percentage points in science (from 59% to 54%), and 6 percentage points in mathematics (from 47% to 41%), in line with the median decline across systems in each of these subjects. On achievement, England (EN) was the only other strong riser in mathematics by 10 percentage points (from 25% to 35%), but did not rise significantly in science (from 45% to 48%). On attitude, as we have already seen, England fell very substantially (by 21 percentage points in science, and 25 in mathematics), at or near the extremes of decline. The example of Massachusetts illustrates that such a fall cannot be attributed to some inevitable within-system trade-off between achievement and attitude. Consequently, in both subjects, England surrendered a considerable lead over Massachusetts on attitude, and fell behind on achievement.

SYSTEMIC CHANGE IN TEACHING EFFECTIVENESS: ENGLISH REFORM REAPPRAISED

While internal evaluation of the impact of English educational reform has focused on rising student achievement in public assessment, these TIMSS findings introduce a broader perspective. They suggest that, triangulated against measures of achievement external to the system, reform has led to fundamental improvement in mathematics achievement but only superficial improvement in science achievement narrowly tied to national assessment. This differential effectiveness by subject may reflect the greater emphasis accorded to mathematics in reform efforts, particularly in the primary phase, as well as the origin of the *Strategy's* basic pedagogical model in research on the teaching of mathematics. Across both subjects, however, the TIMSS findings highlight a serious neglect of student attitude in English improvement efforts.

Limitations of English Reform Officially Recognised

On this issue of student attitude, the line taken by official evaluations such as survey reports by the schools inspectorate has changed markedly, prompted perhaps by the bald TIMSS findings. In mathematics, for example, from reporting that “in many of the schools visited, pupils are working more positively on mathematics” while noting that “occasionally... lack of planning leads to poor behaviour... [where there are] insufficient opportunities to for pupils to be involved actively and weaknesses in classroom management” (OfStEd, 2004, pp. 21 & 23), the position has become one of general concern about students’ attitudes:

A remarkable degree of consistency existed in much of what pupils said about their experience of learning mathematics... Many pupils, especially in secondary schools, described a lack of variety, which they found dull. Typically, their lessons concentrated on the acquisition of skills, solution of routine exercises and preparation for tests and examinations. (OfStEd, 2008a, p. 53)

Though not conceding the part that official policies might have played in creating such a state of affairs (Prestage & Perks, 2008), the inspectors’ report points to the impact of an overly reductive teaching approach on student understanding as well as attitude:

The fundamental issue for teachers is how better to develop pupils’ mathematical understanding. Too often, pupils are expected to remember methods, rules and facts without grasping the underpinning concepts, making connections with earlier learning and other topics, and making sense of the mathematics so that they can use it independently. (OfStEd, 2008a, p. 5)

In science too, there are similar observations:

[M]uch teaching paid scant regard to what and how pupils were learning. In many lessons, teachers simply passed on information without any expectation of pupils' direct engagement in the process. The objective appeared to be to get notes into books, and then leave the learning to the pupils. (OfStEd, 2008b, p. 17)

Publication of these inspection commentaries presaged major shifts in government education policy that were introduced or announced over the following year. These shifts flowed from a view that the educational gains achievable through prescription on the *Strategy* model had been largely exhausted. Key policy changes included a revision of the national curriculum to reduce its degree of prescriptiveness; the abolition of compulsory national testing at the end of lower-secondary education; and the abandonment of a centrally-driven school improvement strategy.

Pointers from the Massachusetts Comparison

From the preceding TIMSS analysis, Massachusetts emerges as a system that has been relatively successful, compared to England, both in raising student achievement and in containing declines in student attitude. Like England, Massachusetts has had a relatively longstanding systemic improvement programme based on establishing common professional standards and ambitious achievement targets, backed by extensive professional development and strong accountability mechanisms (Driscoll, 2009; Massachusetts Department of Education, 1999). This reflects an international trend in which more generic strategies of school improvement and professional development increasingly provide the framework for systemic reform efforts (Ruthven, 2008). Such approaches recognise that institutional and organisational factors (at the levels of the educational system as a whole and the individual school and department) play an important part in supporting improvement at scale in classroom approaches to teaching and learning. At the same time, in both of these systemic initiatives, the promotion of particular pedagogical tools and techniques appears to have played an important part in structuring and scaffolding development of teaching practices.

We have seen that the pedagogical model promoted in England was influenced more by earlier research on effective teaching of basic skills than by more recent research addressing development of higher-order thinking. In Massachusetts, by contrast, improvement policy appears to have taken account of the wider US reform movement represented by the curricular and pedagogical *Standards* developed by the National Council of Teachers of Mathematics (1989, 2000) and the National Academy of Sciences (1995); standards influenced by the later body of research. An early report on the Massachusetts state systemic initiative, Partnerships Advancing the Learning of Mathematics and Science (PALMS), talks of the “reform-oriented pedagogy... advocated by PALMS” (Massachusetts Department of Education, 1999; p. 18). This improvement policy resulted in many Massachusetts schools becoming early adopters of “*Standards-based*” courses designed to operationalise these newer models of curriculum and pedagogy aimed at supporting higher-order thinking and learning. One study noted how, as school districts aligned curriculum and teaching practices with the new state frameworks, reform-oriented, *Standards-based* mathematics programs replaced more traditional ones (Riordan & Noyce, 2001). This study found that students in schools where mathematics was taught through such programs performed significantly better on statewide tests than students following more traditional programs in matched comparison schools. Likewise, the PALMS report draws attention to a study showing “a small, but statistically significant, positive effect of reform-oriented

pedagogy... for mathematics at all three grade levels [4, 8 & 10] and for science at grades 4 and 8” (Massachusetts Department of Education, 1999, p. 18).

CURRENT APPROACHES TO SYNTHESISING RESEARCH ON EFFECTIVE TEACHING

There is, then, an accumulating body of new research arising from systemic efforts to mainstream pedagogical approaches which seek to establish more effective connections between teaching and learning. In particular, between the original proposal for the *epiSTEMe* project and its start, several highly relevant syntheses of research on effective mathematics and science teaching became available. Clearly, this updated body of research has potential to guide future improvement efforts, particularly in addressing student attitude as an educational outcome as much as student achievement. Recent syntheses of research on effective teaching in school mathematics and science have employed a range of methods. The account that follows does not aim to be exhaustive but simply to sketch significant current approaches.

Focused Systematic Review

In Britain, “systematic reviews” of pedagogical research have been carried out under the umbrella of the government-supported Evidence for Policy and Practice Initiative (EPPI) (Bennett et al., 2005). Each synthesis is conducted by a core team of researchers, guided by a wider review group containing a range of practitioners, other professionals and policymakers. This guidance is particularly important in framing and focusing the research questions to be addressed by the review. The synthesis itself is conducted according to explicit protocols that underpin the use of systematic procedures to identify relevant research reports, map their contributions through keywording, and assess their quality through in-depth review. These reviews are highly focused: several of the science-education reviews, for example, have examined small group discussion in science teaching in relation to particular types of learning process or outcome (Science Education Review Group, 2009), while topics examined by the mathematics-education reviews include strategies to raise pupils’ motivational effort at upper secondary level, and effective teacher-initiated teacher-pupil dialogue to promote conceptual understanding at upper-primary and lower-secondary levels (Mathematics Education Review Group, 2009). Although the protocols permit all types of study to be considered, this tight focus narrows the range of directly relevant studies, particularly when restricted to the British research base alone. For example, conducting a systematic review of strategies to raise pupils’ motivational effort in secondary mathematics, Kyriacou & Goulding (2006) identified 25 cognate British studies. Of these, however, “relevance of the focus of the study for the review question” was judged to be “low” in 19 cases and “high” in none; likewise, the “appropriateness of design and analysis for the review question” was judged to be “low” in 19 cases and “high” in only one.

Holistic Negotiated Synthesis

In New Zealand, too, there has been government sponsorship for broader research syntheses on effective pedagogy in science (Hipkins et al., 2002), and in mathematics (Anthony & Walshaw, 2007). The approach taken by both syntheses went beyond the conventional research review by virtue of their concerns with informing policy development and building professional consensus. Indeed the later mathematics review followed a protocol, developed in New Zealand, for “best evidence synthesis” of

research studies (Alton-Lee, 2004) through negotiating a collective understanding of effective practice across the multiple constituencies forming a professional community by means of dialogue between a core research team and a wider professional review group (Anthony & Walshaw, 2007). This type of approach to best-evidence synthesis acknowledges that teaching has a value rationality as well as an instrumental one, and that reform is a political process as much as a scientific one.

Taking a sceptical view of simple mechanisms linking teaching practices to learning outcomes, the mathematics synthesis was reluctant to endorse particular teaching techniques, suggesting that effective teaching is better understood at a level of overarching principles. These principles were extracted from commonalities in the educational values of what the review identified as “landmark studies”:

Although diverse in focus, the researchers that we have reported on are committed to teaching that is less about transmission or delivery of new knowledge and more about taking students’ thinking seriously. Their commitment to students’ thinking is underpinned by the following:

- an acknowledgement that all students, irrespective of age, have the capacity to become powerful mathematical learners;
- a commitment to maximise access to mathematics;
- empowerment of all to develop mathematical identities and knowledge;
- holistic development for productive citizenship through mathematics;
- relationships and the connectedness of both people and ideas;
- interpersonal respect and sensitivity;
- fairness and consistency.

(Anthony & Walshaw, 2007, p. 21)

Likewise, the earlier science synthesis identified core principles for effective inclusive teaching of science, shown in Table 6 (Hipkins et al., 2002, pp. 230-231). This concern with principles over techniques generates a coherent system of orienting heuristics for the teaching process, but it does not give priority to the identification of concrete apparatus that might help scaffold pedagogical change and the professional development of teachers (Ruthven, 2005).

Meta-Analysis of Teaching Effects

As we have already seen, there is a substantial tradition of research on effective teaching techniques, predominantly conducted in the United States. Recent years have seen renewed advocacy and continuing conduct of quantitative studies of teaching and learning effects; and of a variety of approaches to synthesising them (Green & Skukauskaitė, 2008). As well as conventional meta-analysis, these include another form of “best evidence synthesis” – very different in approach from its New Zealand namesake – akin to statistical meta-analysis but providing a fuller description of each contributory study (Slavin, 1986, 2008). Such approaches seek to employ systematic and transparent procedures, first to amass available evaluation studies that provide quantitative measures of teaching effectiveness, and then to identify broad trends across these studies. The aim is to provide guidance that is directly actionable and generally applicable. Compared to the forms of review already discussed, the way in which teaching is conceptualised within such syntheses tends to be less nuanced but more concrete. Meta-analysis can be seen as a useful first stage in a process of progressive focusing, allowing promising lines of pedagogical development to be identified, that can then be formulated in a finer-grained way in the light of the other types of review already discussed.

TABLE 6: Principles for effective inclusive science teaching according to the Hipkins et al. synthesis

- The existing ideas and beliefs that learners bring to a lesson are elicited, addressed, and linked to their classroom experiences;
- Science is taught and learned in contexts in which students can make links between their existing knowledge, the classroom experiences, and the science to be learnt;
- The learning is set at an appropriate level of challenge and the development of ideas is clear – the teacher knows the science;
- The purpose(s) for which the learning is being carried out are clear to the students, especially in practical work situations;
- The students are engaged in thinking about the science they are learning during the learning tasks;
- Students’ content knowledge, procedural knowledge, and knowledge about the nature and characteristics of scientific practice are developed together, not separately;
- The students are engaged in thinking about their own and others’ thinking, thereby developing a metacognitive awareness of the basis for their own present thinking, and of the development of their thinking as they learn;
- The teacher models theory/evidence interactions that link conceptual, procedural, and nature-of-science outcomes, and discussion and argumentation are used to critically examine the relationship between these different types of outcomes;
- Key features of the nature of science are made visible to students and they develop a metacognitive awareness of the similarities and differences between their own personal theorising, and scientific theorising;
- Conversations and investigative skills are scaffolded by the teacher, with explicit modelling of the type of discourse/activity that is appropriate and of the type of outcome/product to be achieved;
- The role of models, modelling, metaphor, and analogies in science is made an explicit focus of practical investigation and of discussion; and
- Teachers engage in formative interactions to help students as they learn.

Reproduced by kind permission of the New Zealand Ministry of Education and the authors

SYNTHESISING RESEARCH ON TEACHING EFFECTS TO IDENTIFY PROMISING APPROACHES

Standard literature searches identified three recent syntheses of meta-analytic type that had produced broad findings about the effective teaching of mathematics and/or science to students in mainstream schooling:

- A meta-analysis had sought to identify effective teaching strategies in science from studies conducted in the United States (Schroeder et al., 2007).
- Parallel best-evidence syntheses (of the meta-analytic type) had surveyed the effectiveness of specific mathematics programs at the elementary school level (Slavin & Lake, 2007, 2008), and at the middle and high school levels (Slavin, Lake & Groff, 2008, 2009).
- A meta-analysis of research on effective teaching and learning components had reported, as a by-product, findings specifically related to science and mathematics (Seidel & Shavelson, 2007).

TABLE 7: Key characteristics of recent meta-analytic syntheses of research on effective mathematics and science teaching

Characteristic	Synthesis		
	<i>Schroeder et al.</i>	<i>Slavin et al.</i>	<i>Seidel & Shavelson</i>
Subject area	Science	Mathematics	All subject areas; with separate reports for science and mathematics
Teaching construct	Teaching strategies	Teaching programs	Teaching components
Outcome(s) examined	Achievement only	Achievement only	Cognitive [achievement] Affective [attitude] Learning process
School levels included	Grades K-12	Grades K-12	Grades 1-13
Study locations included	Restricted to US	Unrestricted, but predominantly US	Unrestricted, but predominantly US and Europe
Publication dates covered	1980-2004	1970-2008	1995-2004
Study duration required	Unrestricted	Required to last at least 12 weeks	Unrestricted
Research design(s) accepted	Experimental or quasi-experimental comparison or evaluation	Randomised experimental or matched quasi-experimental comparison	Correlational survey; Experimental or quasi-experimental comparison
Control for prior student characteristics	Studies not required to control for prior student characteristics	Studies required to show pretest group differences below 0.5 SD, permitting suitable control and adjustment	Studies required to control for prior student characteristics
Achievement measures accepted	Unrestricted, but predominantly non-standardised and researcher developed	Required to show no intervention bias, and so predominantly standardised	Unrestricted
Effect sizes included	Relative [comparative designs] Absolute [non-comparative designs]	Relative	Relative
Aggregation method for effect sizes	Weighted mean by sample size	Median [elementary synthesis] Weighted mean by sample size [middle/high synthesis]	Weighted mean by sample size

While all three syntheses examined effects of teaching on student achievement or cognitive outcomes, only Seidel & Shavelson surveyed effects on learning processes and affective outcomes. Each of the synthesising teams employed somewhat different protocols to govern key decisions in the meta-analytic process, as summarised in Table 7. One important difference is that the Schroeder et al. synthesis did not require studies to control for prior student characteristics (which diminishes the conclusiveness of their findings), and accepted single-group experiments for inclusion (with effect size

measured in absolute terms, rather than relative to another group). Another important difference is that, whereas the other two reviews accepted a wide range of achievement measures, Slavin et al. rejected comparisons based on aspects of achievement likely to have received little or no attention in control groups, so that the studies included in their synthesis predominantly employed standardised tests and state assessments. There is also a surprising lack of overlap in the studies included in the three reviews: the most striking illustration of this is that none of the 32 studies included by Schroeder et al. which were eligible, by virtue of publication date, for the Seidel & Shavelson synthesis, featured in the latter⁷. Finally, as we shall see, there are crucial differences in the classification of studies included in both the Slavin et al. and Seidel & Shavelson syntheses.

Schroeder et al. Meta-Analysis of the Effectiveness of Science Teaching Strategies

The Schroeder et al. (2007) meta-analysis sought to identify types of teaching strategy that are effective in improving student achievement in science. The types of science teaching strategy that the organising framework identifies are shown in Table 8.

TABLE 8: Framework of teaching strategy types in the Schroeder et al. synthesis of science teaching studies

Strategy type	Characterisation (taken from original)
Assessment	Teachers change the frequency, purpose, or cognitive levels of testing/evaluation (e.g., providing immediate or explanatory feedback, using diagnostic testing, formative testing, retesting, testing for mastery)
Collaborative Learning	Teachers arrange students in flexible groups to work on various tasks (e.g., conducting lab exercises, inquiry projects, discussions)
Direct Instruction	Teachers deliver information verbally or explicitly guide students through a sequence of tasks (e.g., learning by listening, designing experiments, using a microscope, making measurements)
Enhanced Context	Teachers relate learning to students' previous experiences or knowledge or engage students' interest through relating learning to the students'/school's environment or setting (e.g. using problem-based learning, taking field trips, using the schoolyard for lessons, encouraging reflection)
Enhanced Materials	Teachers modify instructional materials (e.g., rewriting or annotating text materials, tape recording directions, simplifying laboratory apparatus)
Focusing	Teachers alert students to the intent of the lesson or capture their attention (e.g., providing objectives or reinforcing objectives at the middle or closing of the lesson, using advance organizers)
Inquiry	Teachers use student-centered instruction that is less step-by-step and teacher-directed than traditional instruction; students answer scientific research questions by analyzing data (e.g., using guided or facilitated inquiry activities, laboratory inquiries)
Instructional Technology	Teachers use technology to enhance instruction (e.g., using computers, etc., for simulations; modeling abstract concepts and collecting data; showing videos to emphasize a concept; using pictures, photographs or diagrams)
Manipulation	Teachers provide students with opportunities to work or practice with physical objects (e.g., developing skills using manipulatives or apparatus, drawing or constructing something)
Questioning	Teachers vary timing, positioning, or cognitive levels of questions (e.g., increasing wait time, adding pauses at key student-response points, including more high-cognitive-level questions, stopping visual media at key points and asking questions, posing comprehension questions to students at the start of a lesson or assignment)

TABLE 9: Mean effect sizes for teaching strategy types in the Schroeder et al. synthesis of science teaching studies

Strategy type	Mean effect size [Number of studies]
Assessment	0.51 [2]
Collaborative Learning	0.96 [3]
Direct Instruction	– [0]
Enhanced Context	1.48 [6]
Enhanced Materials	0.29 [12]
Focusing	– [0]
Inquiry	0.65 [12]
Instructional Technology	0.48 [15]
Manipulation	0.57 [8]
Questioning	0.74 [3]

For each strategy type (other than the two for which no relevant studies were found: *direct instruction* and *focusing*), Table 9 shows the mean effect size and the number of contributing studies. Bearing in mind that aggregate effect sizes in this synthesis may be inflated (for reasons discussed above), it seems appropriate to take the (weighted) mean effect size across all strategy types as a convenient benchmark: 0.67. It is important to note also that this value is strongly influenced by the high weighting of the *inquiry* category, for which the relevant studies contributed no less than 91% of student numbers (on which weighting was based). Two strategy types place clearly above this overall mean (on the basis that the lower limit of the confidence interval quoted for their effect size exceeds it): *enhanced context* and *collaborative learning*. Two strategy types place clearly below the overall mean (on the basis that the upper limit of the confidence interval quoted for their effect size falls below it): *enhanced materials* and *instructional technology*; and one further strategy type comes very close to doing so: *manipulation*. Finally, the remaining two strategy types have relatively wide confidence intervals for effect size because the relevant studies were small scale in terms of student numbers: *questioning*, with a mean effect size above the overall mean; and *assessment*, with a mean effect size below.

The findings of this synthesis point, then, to *enhanced context* and *collaborative learning* as particularly promising strategy types in terms of their impact on achievement outcomes from science teaching. In addition, the effectiveness of the *inquiry* strategy type is supported by research involving substantial numbers of students. There is more tentative support for the *questioning* strategy type in view of the more slender research base within the review (both in terms of number and size of studies).

Slavin et al. Best Evidence Synthesis of the Effectiveness of Mathematics Teaching Programs

The primary aim of the Slavin et al. best evidence syntheses was to grade mathematics teaching programs according to the scale, quality and strength of evidence of their effectiveness in promoting student achievement. A secondary aim was to draw wider conclusions about common characteristics of successful programs, through examining the overall effectiveness of cognate categories of programs. The two syntheses employ similar organising frameworks for classifying programs⁸, as summarised in Table 10.

**TABLE 10: Framework of teaching program types
in the Slavin et al. syntheses of mathematics teaching studies**

Program type	Subtype	Characterisation (adapted from originals)
Mathematics Curricula		Programs based on use of textbook series
	Back-to-basics	Textbooks that emphasise building students' confidence and skill in mathematics through a step-by-step approach to standard computations and problems
	NSF-funded	Textbooks based on the NCTM Standards that emphasise problem solving, alternative solutions, and conceptual understanding
	Traditional	Textbooks that provide a more traditional balance among computations, concepts, and problem solving
Computer Assisted Instruction		Programs based on use of computer-based tutorial and learning management systems
Instructional Process		Programs based on professional development in particular instructional techniques
	Classroom Management & Motivation	Techniques to make effective use of lesson time, and to enhance student motivation
	Cooperative Learning	Techniques in which students work in pairs or small groups to help each other master academic content.
	Cooperative/Individualized Learning	Techniques that combine cooperative learning with individualised instruction
	Direct Instruction	Techniques that emphasise a structured, step-by-step approach to key concepts
	Individualized Instruction	Techniques that diagnose individual students' strengths and weaknesses and give them appropriate material to meet learning objectives with teacher and peer support
	Mastery Learning	Techniques in which students are taught to well-defined standards, formatively assessed, and given corrective instruction if needed
	Mathematics Content	Techniques focused on helping children learn through building on their intuitive knowledge of mathematics
	Metacognitive Strategy Instruction	Techniques in which students are taught to ask themselves aloud questions of comprehension, connections and similarities/ differences with previous problems, appropriate strategies, and reflection

The findings for these program types are shown for elementary schools (in Table 11) and middle/high schools (Table 12). At both levels, those programs judged to have the strongest evidence of effectiveness came from the *instructional process* category. The majority of programs were judged to lack evidence of effectiveness by virtue either of the absence of adequate studies, or of the weak findings of such studies.

Within the category of *mathematics curricula*, the specific programs for which evidence of effectiveness was found include mathematics curricula of all three types, but all with relatively modest effects. While some textbook series proved a little more effective in raising student achievement on conventional standardised tests than others, such effects were not strong, and they appear to be largely unrelated to the adoption of a traditional, back-to-basics, or *Standards*-based approach as such. The small aggregate effect sizes for each of the three subcategories suggest that no type of mathematics curriculum is, in general, particularly effective in the terms adopted for this synthesis.

TABLE 11: Median effect sizes for teaching program types in the Slavin & Groff synthesis of elementary school mathematics teaching studies

Program type	Subtype	Median effect size [number of studies]
Mathematics Curricula		0.10 [13]
	Back-to-basics	0.02 [1]
	NSF-funded	0.12 [6]
	Traditional	0.10 [7]
Computer Assisted Instruction		0.19 [38]
Instructional Process	Classroom Management & Motivation	0.41 [6]
	Cooperative Learning	0.29 [9]
	Cooperative/Individualized Learning	0.24 [6]
	Direct Instruction	0.46 [4]
	Mastery Learning	0.22 [5]
	Mathematics Content	0.28 [2]

TABLE 12: Mean effect sizes for teaching program types in the Slavin et al. synthesis of middle/high school mathematics teaching studies

Program type	Subtype	Mean effect size [number of studies]
Mathematics Curricula		0.03 [40]
	Back-to-basics	0.14 [11]
	NSF-funded	0.00 [26]
	Traditional	0.13 [3]
Computer Assisted Instruction		0.08 [40]
Instructional Process	Cooperative Learning	0.42 [8]
	Individualized Instruction	0.36 [2]
	Mastery Learning	-0.05 [6]
	Metacognitive Strategy Instruction	0.31 [2]

Limitations of space mean that the findings for *computer-assisted instruction* cannot be discussed in depth here. By way of summary however, at both levels of schooling, aggregate effect sizes were intermediate between those for mathematics curricula and instructional processes, but closer to the former.

The range of *instructional process* programs identified was somewhat different at the two levels of schooling. However, focusing on the specific programs for which strong evidence of effectiveness was found, 4 out of 5 at the elementary level, and 2 out of 2 at the middle/high level were categorised as *cooperative learning*. This is reflected also in the relatively large aggregate effect size for this category at both levels. At middle/high level Slavin and colleagues also single out *metacognitive strategy instruction* as another promising approach, and *individualized instruction* (although here there were no supporting studies after the early 1970s). Likewise, at elementary level,

they highlight the effectiveness of *classroom management and motivation* programs (although the Missouri Mathematics Program that figures prominently in this category would perhaps be better labelled in its own broader terms of *active teaching*) and what appear (from comparison of the fuller descriptions provided) to be cognate *direct instruction* programs.

It is important to note that the *instructional process* category requires the involvement of teachers not just in using a particular instructional process, but in following professional development that focuses on it. Thus the relative success of *instructional process* programs may testify as much to the role of professional development in supporting improvement as to the effectiveness of the particular instructional processes involved. For example, many of what were classed by Slavin et al. as *NSF-funded mathematics curricula* have a cooperative learning aspect; had these programs also been included in the *cooperative learning* subcategory, its aggregate effect size would have been very much lower. Correspondingly, studies of *Standards-based mathematics courses in classroom use* have reported weak implementation of the cooperative learning aspect (Arbaugh et al., 2006; Lloyd, 2008), but found that successful implementation of it is strongly associated with student achievement (Schoen et al., 2003). Equally, the relevant EPPI systematic reviews conclude that the effectiveness of cooperative learning through small group discussion depends on both teachers and students receiving explicit training in the skills associated with sustaining discussion and developing arguments (Bennett et al., 2010).

Seidel & Shavelson Meta-Analysis of the Effectiveness of Teaching Components

The main focus of the work conducted by Seidel & Shavelson (2007) was on the way in which the results of meta-analysis are shaped by the overarching framework chosen to organise the multitude of teaching variables featuring in the studies under review. Seidel & Shavelson found that a new framework based on cognitive models of teaching and learning proved more sensitive in detecting components of effective teaching than a more traditional framework focusing on aspects of teaching method. This new framework reflects important shifts that have taken place in the way in which teaching processes are conceptualised within educational psychology and cognitive science, linking them more closely to the learning activity of students. Likewise, the conceptual framework that Seidel & Shavelson employed to classify learning outcomes (or teaching products) reflects greater attention to the learning processes employed by students, and to affective outcomes alongside cognitive ones. The core of this new framework comprised those teaching components conjectured to lead most directly to student learning. These are shown in Table 13, each accompanied by a brief characterisation.

In the meta-analysis that Seidel & Shavelson conducted using this framework one teaching component emerged as particularly effective: *domain-specific information processing*. Moreover, this component proved effective “regardless of domain (reading, mathematics, science), stage of schooling (elementary, secondary), or type of learning outcome (learning processes, motivational-affective, cognitive)” (Seidel & Shavelson, 2007, p. 483). In the course of examining the robustness and generality of this finding, Seidel & Shavelson reported separate results for science teaching (Table 14) and mathematics teaching (Table 15)⁹. These tables highlight some limitations of the evidence base available to them, particularly the under-researching of learning processes and affective outcomes as against cognitive achievement: this is shown not just by the empty cells¹⁰ but by the small numbers of replications in other cells under the learning processes and affective outcomes heads. However, in line with the broader

TABLE 13: Framework of core teaching components in the Seidel & Shavelson synthesis of teaching studies

Component type	Characterisation (adapted from original)
Goal setting and orientation	Teaching components that clarify learning goals and orient students towards learning to achieve them. Establishing clear learning goals; activating student pre knowledge; using anchors and contexts in explaining learning contents
Execution of learning activities	
Social interactions/ direct experiences	Teaching components that support social interactions between students and provide direct experiences for students. Cooperative learning; student hands-on activities; student discussions; use of a variety of teaching methods
Basic information processing	Teaching components that facilitate the basic processing of information. Cognitive activation; active cognitive student engagement; use of a high language level to engage students in higher order thinking; think-aloud training
Domain-specific information processing	Teaching components that provide domain-specific opportunities for processing content information, such as those involved in mathematical problem solving, scientific inquiry, or specific reading and writing strategies
Evaluation of learning	Teaching components that aim to assess student progress toward learning goals
Regulation/monitoring	Teaching components that provide support through feedback and monitoring, and support students' self-regulation of learning

findings, *domain-specific information processing* is the most promising teaching component, found to be effective in terms of cognitive outcomes in both science and mathematics, in terms of affective outcomes in science (with insufficient evidence available in mathematics), and in terms of learning processes in mathematics (with insufficient evidence available in science).

There is less clear-cut promise for the teaching component of *social interactions/direct experiences*. In science, this is found to benefit affective outcomes, and, to a lesser degree, learning processes, but not cognitive outcomes. In mathematics, the assembled evidence indicates that this teaching component is not effective in relation to either affective or cognitive outcomes. Some useful clarification is provided by the findings (from all studies, not just those in mathematics and science) for *cooperative learning* which figures as a teaching component under the more traditional framework examined by Seidel & Shavelson (Table 16). This shows that, for cognitive outcomes, the findings under *social interactions/direct experiences* in both subjects are in line with the more general findings under *cooperative learning*. For learning processes and affective outcomes, there is a similar pattern in science. However, the findings from a small number of replications in relation to affective outcomes in mathematics diverge from this wider pattern.

TABLE 14: Mean effect sizes for teaching component types in the Seidel & Shavelson synthesis of science teaching studies

Component type	Learning effect		
	Mean effect size [Number of replications]		
	Learning processes	Affective outcomes	Cognitive outcomes
Goal setting and orientation	.12 [21]		.22 [6]
Social interactions/direct experiences	.20 [13]	.41 [14]	.00 [35]
Basic information processing	.06 [23]		.06 [15]
Domain-specific information processing		.35 [7]	.63 [28]
Evaluation of learning			
Regulation/monitoring	.06 [15]		

TABLE 15: Mean effect sizes for teaching component types in the Seidel & Shavelson synthesis of mathematics teaching studies

Component type	Learning effect		
	mean effect size [number of replications]		
	Learning processes	Affective outcomes	Cognitive outcomes
Goal setting and orientation	-.06 [6]	.16 [13]	.04 [64]
Social interactions/direct experiences		.02 [9]	-.04 [42]
Basic information processing	.06 [7]	.16 [35]	.02 [102]
Domain-specific information processing	.41 [10]		.37 [22]
Evaluation of learning			.06 [34]
Regulation/monitoring		.20 [22]	-.08 [50]

TABLE 16: Mean effect sizes for cooperative learning in the Seidel & Shavelson synthesis of teaching studies

Component type	Learning effect		
	mean effect size [number of replications]		
	Learning processes	Affective outcomes	Cognitive outcomes
Cooperative learning	.22 [12]	.26 [24]	.00 [149]

Triangulating and Summating the Meta-Analyses of Effective Teaching

The Seidel & Shavelson meta-analysis provides the most convenient anchor for drawing together the findings of the three syntheses because it covers both subject areas. Moreover, it has the advantage of attending to learning processes and affective outcomes as well as to cognitive achievement. In this section, then, its main findings will be triangulated and extended by comparing them with those of the other subject-specific syntheses. This integration is not straightforward, however, because of the differing conceptual frameworks used in the three syntheses.

Domain-specific inquiry

The most broad and clear-cut finding from the Seidel & Shavelson synthesis is the effectiveness of teaching components that focus on *domain-specific information processing* with respect to all types of outcome for which evidence is available. A reasonably close match for this component in the Schroeder et al. framework appears to be the *inquiry* strategies found to be effective in terms of student achievement in science. There appears, however, to be no ready counterpart for *domain-specific information processing* within the Slavin et al. framework, although *NSF-funded mathematics curricula* would, as one of several characteristic features, involve students in problem solving akin to the Seidel & Shavelson and Schroeder et al. descriptors. In the Slavin et al. synthesis, of course, such curriculum programs were not found to be particularly effective in terms of the narrower range of measures of achievement accepted by that review. However, because a primary aim of *Standards-based* programs is also to address non-traditional achievement outcomes, they have been designed to maintain rather than improve student performance on traditional measures (Schoenfeld, 2006). As the reviewers themselves note, and others have emphasised (Confrey, 2006; Schoenfeld, 2006), evaluation restricted to traditional measures will miss wider effects:

The research... is at least comforting in showing that reform-oriented curricula are no less effective than traditional curricula on traditional measures, so their contribution to non-traditional outcomes does not detract from traditional ones.

(Slavin et al., 2009, pp. 886-7)

Another important factor here appears to be crucial differences in classification between the reviews. For example, Slavin et al. identified IMPROVE as a program with strong evidence of effectiveness, and it also features in the Seidel & Shavelson review. However, it is classified as *cooperative learning* by Slavin et al., but as *domain-specific information processing* by Seidel & Shavelson. In discussing this particular program, Slavin and colleagues note that it also has a metacognitive aspect, and they categorise a related program as *metacognitive strategy instruction*, another subcategory found to be relatively effective. Thus it becomes clear that some effective interventions classed by Seidel & Shavelson as *domain specific information processing* were placed by Slavin et al. within apparently quite different *instructional process* subcategories.

In summary, then, the available evidence suggests that a teaching component that might be designated *domain-specific inquiry* (to combine the Seidel & Shavelson and Schroeder et al. terms) is effective in both subjects across all types of outcome, although there remains scope for further investigation of specific aspects on which evidence is currently lacking.

Cooperative groupwork

A more muted finding from the Seidel & Shavelson synthesis is the effectiveness of *social interactions/direct experiences* with respect to learning processes and affective outcomes, but only in science. Many of the studies classed as *social interactions/direct experiences* in the more modern framework used by Seidel & Shavelson were classed as *cooperative learning* in their more traditional framework, cognate to *cooperative learning* in the Slavin et al. framework and *collaborative learning* in Schroeder et al. However, contrary to Seidel & Shavelson, the other syntheses found this to be a relatively effective teaching component in promoting student achievement. In the first instance, these conflicting findings seem to arise from differences in selection and classification between the syntheses. To take a key example of differences in selection, Slavin et al. single out Student Teams – Achievement Divisions (STAD) as a program with strong evidence of effectiveness at both school levels on the basis of 8 studies; but

while 4 of these studies appeared during the period covered by the Seidel & Shavelson review, none were included in it. The IMPROVE program has already provided another key example of a difference in classification, reflecting the way in which *cooperative* or *collaborative learning* can be combined with other types of teaching component, notably Seidel & Shavelson's *domain-specific information processing* and Schroeder et al.'s *inquiry*. At a more fundamental level, too, as the contrast between the STAD and IMPROVE programs illustrates, cooperative or collaborative learning covers a range of approaches somewhat different in form and rationale; varying, for example, in the degree to which they emphasise motivational factors linked to the goal and reward structures under which group members operate, as against cognitive factors linked to the intellectual character and quality of interactions between group members.

In summary, then, the available evidence suggests that a teaching component which might be designated *cooperative groupwork* (to combine commonly used terms) is effective in science in relation to learning processes and affective outcomes. In mathematics, however, this component is unproven in relation to learning processes and neutral in relation to affective outcomes. In both subjects, the syntheses conflict in their conclusions about effectiveness in relation to achievement outcomes. Seidel & Shavelson find a null effect in both subjects, whereas both Schroeder et al. in science and Slavin et al. in mathematics report relatively strong positive effects. This is a crucial discrepancy which requires further analysis.

Contextual orientation

Seidel & Shavelson's category of *goal setting and orientation* merits some attention. In science, it was found to be moderately effective in relation to learning processes and cognitive outcomes, with evidence lacking on affective outcomes. There appears to be overlap here with what Schroeder et al. found to be highly effective *enhanced context* strategies in science in which learning is related to students' prior experience or knowledge. However, this seems closer to *goal orientation* – activating student pre-knowledge, and supporting explanation through using anchors and contexts – than to *goal setting* – establishing clear learning goals – which corresponds rather to Schroeder et al.'s *focusing*. Indeed, the fact that Schroeder et al. found no studies of *focusing* in science suggests that *goal orientation* has been more favoured there than *goal setting*. In mathematics, *goal setting and orientation* was found to be moderately effective in relation to affective outcomes, but not learning processes or cognitive outcomes. In mathematics, triangulation is not possible because the Slavin et al. synthesis has no category corresponding closely enough either to Seidel & Shavelson's *goal setting and orientation*, or to Schroeder et al.'s *focusing* or *enhanced context*.

In summary, then, a more speculative combination of findings from Seidel & Shavelson and Schroeder et al. points to the effectiveness of what might be termed a teaching component of *contextual orientation* in promoting learning processes and achievement outcomes in science. This deserves to be researched more fully in relation to mathematics.

Other components

From the studies of *evaluation of learning* that Seidel & Shavelson found, they were able to conclude only that this teaching component is not particularly effective in promoting cognitive outcomes in mathematics. Equally, Schroeder et al. found that *assessment* strategies were less effective than average in promoting student achievement in science. In Slavin et al.'s framework there appears to be no category close enough to support triangulation. In summary, then, these syntheses do not provide support for the

effectiveness of what might be termed a teaching component of *learning assessment*. The apparent discrepancy of these findings with a (more focused but not subject-specific) synthesis that has reported the effectiveness of *formative assessment* strategies (Black & Wiliam, 1998) may be explicable in terms of there being little formative aspect to the evaluation and assessment strategies featuring in the studies reviewed by Seidel & Shavelson and Schroeder et al. Indeed, effective types of *formative assessment* may be closer to Slavin et al.'s *metacognitive strategy instruction* or Seidel & Shavelson's *regulation/monitoring* (Anthony, 1996; Ruthven, 2002). There, however, these syntheses arrive at conflicting findings in relation to mathematics achievement, and lack evidence in science. This is an area that deserves further analysis and investigation.

Finally, turning to categories from the other syntheses for which there are no close parallels in the Seidel & Shavelson framework, perhaps the most important are the cognate *classroom management and motivation* and *direct instruction* programs that Slavin et al. found to be effective in mathematics, termed *active teaching* by influential researchers in this area. The Schroeder et al. synthesis reported no studies of *direct instruction* in science, but did identify enhanced teacher *questioning* as a relatively effective strategy, apparently in a context of interactive whole-class teaching. Overall, then, the syntheses provide evidence of the effectiveness of what might be termed an *active teaching* component in promoting achievement in mathematics. This type of pedagogical approach deserves to be researched more fully, particularly in relation to science.

CONCLUSION

This paper has shown, first, how the TIMSS international study series can be used to construct a framework for cross-system and between-subject comparison capable of triangulating findings from national assessment, and of illuminating trends in student achievement and attitude. Applied to the case of English lower-secondary schooling, such triangulation has indicated that, contrary to the more optimistic picture from national assessment, fundamental gains in student achievement have taken place only in mathematics, and only in response to the second wave of reform associated with the national *Strategy*. Likewise, contrary to the relative subject profiles in national assessment, the TIMSS performance of English students has been stronger in science than in mathematics. Finally, against an international trend for student attitude towards mathematics and science to decline, the TIMSS comparisons have shown recent falls in both subjects to be exceptionally severe amongst English students.

Second, this paper has shown how systematic research syntheses, specifically those adopting meta-analytic approaches, can be used to identify promising lines of pedagogical development. While it has become clear that the findings of any systematic synthesis depend to some degree on crucial features of the particular system employed – notably the conceptual framework embraced, the criteria adopted for screening studies, and the strategies used to identify relevant studies within a sparse and scattered literature –, through triangulation of such syntheses it has proved possible to establish reasonably robust findings about the relative effectiveness of four teaching components: *domain-specific inquiry* in relation to student achievement in both subjects, student attitude in science, and learning processes in mathematics; *cooperative groupwork* in relation to learning and attitude in science; *contextual orientation* in relation to achievement in science; and *active teaching* in relation to achievement in mathematics. Equally, discrepancies between findings or insufficiencies of evidence have highlighted a number of issues particularly deserving further analysis and investigation: the impact

of *cooperative groupwork* on achievement outcomes; the impact of differing forms of *learning assessment* on both achievement and attitude outcomes; the impact of *contextual orientation* on outcomes in mathematics; the impact of *active teaching* on outcomes in science.

These two lines of enquiry converge in the finding that reform with an emphasis on *active teaching* in English schools has been associated with a fundamental gain in student achievement in mathematics but not in science. However, this reform has also been associated with a severe decline in student attitudes to both subjects. It may be, of course, that the recent relaxation of national testing requirements will reduce the heavy emphasis on test preparation, and that this will improve student attitudes. Focusing, though, on the longer-term and more far-reaching pedagogical development still required at lower-secondary level in England, a plausible way forward is much clearer in science than in mathematics.

In science, the research syntheses indicate that emphasis on *domain-specific inquiry* should be beneficial for affective outcomes and cognitive achievement, *cooperative groupwork* for affective outcomes and learning processes, and *contextual orientation* for learning processes and cognitive outcomes. In mathematics, however, while some teaching components were found to be moderately effective in relation to affective outcomes, these proved either ineffective or unproven in relation to learning processes and cognitive outcomes. *Domain-specific inquiry* emerges as the most promising, on the grounds of its proven effectiveness in relation to learning processes and cognitive outcomes in mathematics (with its impact on affective outcomes under-investigated and so unproven). From the Slavin et al. synthesis there is also support for the effectiveness of *active teaching* in relation to achievement in mathematics (although evidence is lacking on its impact on attitudes).

In the English educational system, the recent policy shifts described earlier have created conditions under which schools and teachers are now less constrained in the kinds of pedagogical development that they can envisage undertaking. Nevertheless, few are well placed to exploit this new flexibility after 20 years of closely regulated centralised prescription. The approach being taken by the *epiSTEMe* project, designing for implementation at scale within the English educational system, is what might best be described as one of ‘redesign research’, working within the established curricular framework and taking account of the existing professional practice from which any development must start. The project is developing a pedagogical model running across mathematics and science that establishes stronger links between the subjects (Ruthven et al., 2010). Informed by a range of relevant theoretical perspectives and by previous research and development, the proposed pedagogical model emphasises components of *domain-specific inquiry*, *contextual orientation*, and *cooperative groupwork* supported by a form of *active teaching* incorporating whole-class discussion.

The research syntheses analysed in this paper suggest that this type of configuration may be better suited to science education than mathematics. In mathematics, of course, configurations of this type have been employed in many of the *Standards*-based curricula developed in the United States, which played a part in the school improvement efforts that produced a strong Massachusetts performance in TIMSS. Equally, the expanded formulation for English reform originally offered by Reynolds & Muijs suggests that development of higher-order thinking in mathematics calls for incorporation of elements both of *domain-specific inquiry* and *cooperative groupwork* to complement *active teaching*. Given the influence of the national *Strategy*, mathematics teaching in English schools can be expected to already emphasise the *active teaching* known to be effective in relation to more traditional outcomes; hence,

introducing, or strengthening, the complementary components emphasised in the *epiSTEMe* model represents a plausible strategy for producing broader improvements in achievement and attitude. Such improvements can be expected to depend, of course, not on the apparatus alone, but on the degree to which it can serve as a vehicle for practical expression of pedagogical heuristics of the type identified in more fine-grained and subject-specific syntheses (e.g. Hipkins et al., 2002, as shown in Table 6), and the professional learning required to achieve this.

NOTES

¹ Within the United Kingdom of Great Britain and Northern Ireland, there are four educational systems: those of England (much the largest), Wales and Northern Ireland (both originally modelled on the English system but increasingly independent of it), and Scotland (historically independent). Nevertheless, there are strong parallels between the systems, particularly at the early-secondary level on which the *epiSTEMe* project is focusing.

² Expert review of the treatment of concept, context and format in TIMSS items has found their likely familiarity to English students to be rather greater in mathematics than in science (Ruddock et al., 2006).

³ The intermediate cohort was not surveyed by TIMSS at Grade 4 level in 1999, and did not complete the main attitude-to-subject measures at Grade 8 level in 2003. There were also some sampling weaknesses at Grade 8 level: England came close to satisfying guidelines for sample participation rates only after replacement schools were included. However, the English national TIMSS report (Ruddock et al., 2004; p. 5) indicates that it proved possible to reweight the data satisfactorily (using schools' relative performance in national tests) to ensure that the TIMSS findings were representative. This cohort has been included in the analysis only where it can provide crucial clarification of trends.

⁴ As well as England (EN), these 18 systems comprise British Columbia (BC), Chinese Taipei (TW), Hong Kong (HK), Iran (IR), Israel (IL), Italy (IT), Japan (JP), Jordan (JO), (South) Korea (KR), Malaysia (MY), Massachusetts (MA), Ontario (ON), Quebec (QC), Singapore (SG), Thailand (TH), Tunisia (TN), United States (US). Although some data from Massachusetts is included as part of that for the United States, the portion is so small, and state-level education policies sufficiently distinctive, that both entities can reasonably be included as systems within the analysis. Likewise, the relative independence of the Canadian provinces in educational matters justifies their treatment as distinct systems.

⁵ However, smaller but still substantial improvements at Grade 4 level in Hong Kong and Ontario did not feed through to Grade 8 level. Such patterns suggest that there is no straightforward relationship between changes at the two levels.

⁶ The stronger alignment between national and TIMSS tests in mathematics than in science (as noted in 2 above) needs to be borne in mind. However, this did not lead to a subject-differentiated pattern of change in performance between 1999 and 2003; over that period rising performance in national tests in mathematics was not paralleled in TIMSS.

⁷ Studies cited in Schroeder et al. (2007) that were published between 1995 and 2004 were checked against studies cited in Seidel & Shavelson (2007). Further checks were then conducted: for example, Schroeder et al. cite 5 studies in the *Journal of Research*

in Science Teaching that appeared during this period, and Seidel & Shavelson 2, but there is no overlap between them.

⁸ Two further types of *instructional process* program which do not focus on core classroom instruction have been omitted: *comprehensive school reform*, and *supplemental*.

⁹ In the interests of comparability with the other syntheses discussed in this paper, the effect sizes are given here in terms of the metric associated with Cohen's *d*, converted from that associated with Fisher's *Z* (Seidel & Shavelson, 2007, p. 471).

¹⁰ Seidel & Shavelson only report aggregate effects sizes based on five or more replications; hence an empty cell indicates four or fewer replications, not necessarily none.

ACKNOWLEDGEMENTS

The *epiSTEMe* project is supported by the UK Economic and Social Research Council (RES-179-25-0003). An earlier version of the TIMSS analysis was presented at the British Congress of Mathematics Education (Ruthven, 2010).

REFERENCES

- Alton-Lee, A. (2004). *Guidelines For Generating a Best Evidence Synthesis Iteration*. Wellington NZ: Ministry of Education.
- Anthony, G. (1996). Active learning in a constructivist framework. *Educational Studies in Mathematics*, 31(4), 349-369.
- Anthony, G. & Walshaw, M. (2007). *Effective Pedagogy in Mathematics/Pàngarau: Best Evidence Synthesis Iteration*. Wellington NZ: Ministry of Education.
- Arbaugh, F., Lannin, J., Jones, D. & Park-Rogers, M. (2006). Examining instructional practices in *Core-Plus* lessons: implications for professional development. *Journal of Mathematics Teacher Education*, 9(6), 517-550.
- 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.
- Black, P. & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education*, 5(1), 7-74.
- Brown, M., Askew, M., Baker, D., Denvir, H. & Millett, A. (1998). Is the National Numeracy Strategy research-based? *British Journal of Educational Studies*, 46(4), 362-385.
- Brown, M., Askew, M., Millett, A. & Rhodes, V. (2003). The key role of educational research in the development and evaluation of the *National Numeracy Strategy*. *British Educational Research Journal*, 29(5), 655-672.
- Confrey, J. (2006). Comparing and contrasting the *National Research Council* report on evaluating curricular effectiveness with the *What Works Clearinghouse* approach. *Educational Evaluation and Policy Analysis*, 28(3), 195-213.
- Department for Children, Schools and Families [DCSF] (2008a). National Curriculum Assessments at Key Stage 3 in England, 2006/07 (Revised). *Statistical First Release 06/2008*. London: DCSF.
- Department for Education and Employment [DfEE] (1998b). *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>
- Economic and Social Research Council [ESRC] (2006). *Targeted Research Initiative on Science and Mathematics Education: Call for Outline Research Proposals*. Swindon: ESRC.
- Good, T.L., Grouws, D.A. & Ebmeier, H. (1983). *Active mathematics teaching*. New York: Longman.

- Green, J. & Skukauskaitė, A. (2008). Becoming critical readers: Issues in transparency, representation, and warranting of claims. *Educational Researcher* 37(1), 30-40.
- 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.
- Lloyd, G. M. (2008). Teaching mathematics with a new curriculum: Changes to classroom organization and interactions. *Mathematical Thinking and Learning*, 10(2), 163-195
- 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.
- Office for Standards in Education [OfStEd] (2004). *The Key Stage 3 Strategy: evaluation of the third year*. London: OfStEd.
- Office for Standards in Education [OfStEd] (2008a). *Mathematics: understanding the score*. London: OfStEd.
- Office for Standards in Education [OfStEd] (2008b). *Success in science*. London: OfStEd.
- Prestage, S. & Perks, P. (2008). HMI Ofsted report for Mathematics 2008 or why teenagers are maths dunces. *Proceedings of the British Society for Research into Learning Mathematics*, 28(3), 96-101.
- 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.
- Ruddock, G., Clausen-May, T., Purple, C. & Ager, R. (2006). *Validation Study of the PISA 2000, PISA 2003 and TIMSS 2003 International Studies of Pupil Attainment* (DfES Research Report 772). London: DfES.
- Ruddock, G., Sturman, L., Schagen, I., Styles, B., Gnaldi, M., & Vappula, H. (2004). *Where England stands in the Trends in International Mathematics and Science Study (TIMSS) 2003: National report for England*. Slough: National Foundation for Educational Research.

- Ruthven, K. (2002). Assessment in mathematics education. In L. Haggarty (Ed.) *Teaching mathematics in secondary schools* (pp. 176-191). London: Routledge-Falmer.
- Ruthven, K. (2005). Improving the development and warranting of good practice in teaching. *Cambridge Journal of Education*, 35(3), 407-426.
- Ruthven, K. (2008). Reflexivity, effectiveness, and the interaction of researcher and practitioner worlds. In P. Clarkson & N. Presmeg (Eds.), *Critical Issues in Mathematics Education* (pp. 213-228). New York: Springer.
- Ruthven, K. (2010). Using longitudinal, cross-system and between-subject analysis of the TIMSS study series to calibrate the performance of lower-secondary mathematics education in England. *Proceedings of the 7th. British Congress of Mathematics Education [BCME 7]*, published as *Proceedings of the British Society for Research into Learning Mathematics* 30(1), 183-190.
- Ruthven, K., Howe, C., Mercer, N., Taber, K. Luthman, S., Hofmann, R. & Riga, F. (2010). Effecting Principled Improvement in STEM Education. *Proceedings of the 7th. British Congress of Mathematics Education [BCME 7]*, published as *Proceedings of the British Society for Research into Learning Mathematics*, 30(1), 191-198.
- Schoen, H., Cebulla, K., Finn, K. & Fi, C. (2003). Teacher variables that relate to student achievement when using a Standards-based curriculum. *Journal for Research in Mathematics Education*, 34(3), 228-259.
- Schoenfeld, A. (2006). What doesn't work: The challenge and failure of the *What Works Clearinghouse* to conduct meaningful reviews of studies of mathematics curricula. *Educational Researcher*, 35(2), 13-21.
- 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. (2007). *Effective programs in elementary mathematics*. Version 1.2. Accessed 6 November 2009, http://www.bestevidence.org/math/elem/elem_math.htm
- Slavin, R. & Lake, C. (2008). Effective programs in elementary mathematics. *Review of Educational Research*, 78(3), 427-515.
- Slavin, R., Lake, C. & Groff, C. (2008). *Effective programs in middle and high school mathematics*. Version 1.4. Accessed 26 July 2009, http://www.bestevidence.org/math/mhs/mhs_math.htm
- 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.