This paper examines the evolving state of research synthesis to inform pedagogical improvement in school mathematics and science. It summarises and extends the analysis published as a section of an earlier journal paper (Ruthven 2011) which provides the supporting detail for which there is not space in this short conference paper.

The main concern of this paper is to provide a critical appreciation and comparison of three distinctive approaches that have been employed in recent syntheses of research on pedagogy in school mathematics and science, and then to triangulate their findings to identify key agreements, anomalies and absences.
This analysis was undertaken with a view to identifying what guidance these syntheses can offer efforts to support pedagogical improvement, such as the *epiSTEMe* project which I am currently leading. This project aims to develop and analyse a research-informed pedagogical intervention, suited to implementation at scale in the English educational system. It forms part of a national research initiative which is investigating ways of tackling participation and achievement gaps in these areas.

**The *epiSTEMe* project**

- Aims to develop and analyse a research-informed pedagogical intervention in lower-secondary physical science and mathematics, suited to implementation at scale in the English educational system.
- Forms part of the ESRC’s Targeted Initiative on Science and Mathematics Education [TISME] which is investigating ways of tackling participation and achievement gaps in these areas.
Recent syntheses of research on effective pedagogy in school mathematics and science

- **Systematic reviews carried out in the UK**
  - under the aegis of the Evidence for Policy and Practice Information [EPPI] and Co-ordinating Centre, and either the Department for Education or the Training and Development Agency for Schools
  - by teams based at the University of York

- **Research syntheses, particularly a best evidence synthesis iteration [BESI], carried out in NZ**
  - under the aegis of the Ministry of Education
  - by teams based at Massey and Waikato Universities

- **Meta-analytic syntheses carried out in the US**
  - by teams based at Johns Hopkins, Texas A&M, and Stanford Universities

This slide provides basic information about the three recent bodies of research synthesis to be examined, notably their organisational sponsors and the university affiliations of the teams undertaking them.

These three bodies of synthesis result from the application of somewhat different approaches. In particular, the EPPI systematic review and the NZ best-evidence synthesis iteration both originate in concerns, on the one hand, about the potential for bias in the conventional connoisseurial narrative review, and, on the other hand, about the restricted scope of the statistically-driven process-product meta-analysis.

With this in mind, the next few slides will sketch these two relatively novel approaches, and identify some key issues that they raise.
The EPPI systematic reviews in mathematics education (Mathematics Education Review Group 2009) and science education (Science Education Review Group 2009) have been carried out by distinct teams following the same review protocol.

The titles of indicative publications from each team are shown in the left-hand panel on the slide.

The main elements of the deliberately systematic and reflexive review process are shown in the right-hand panel: note the shift in emphasis from the systematic to the reflexive as the steps in the process call for increasing levels of judgement and inference.
As the very titles of the reviews suggest, they are tightly focused and specified. Indeed, the review process provides for further tightening of focus at a later stage of in-depth review.

By way of example, an initial review of research on small-group discussions in science teaching with students aged 11-18 led to a more specific focus on impact on students’ understanding of evidence in science. Equally, the research studies had to meet rather stringent criteria for inclusion in the review.

The review teams have themselves identified two consequential problems: first, the typically small number of research studies directly meeting all criteria limits the available evidence base; and, second, an important degree of idiosyncrasy in many of the suitable studies makes synthesis more challenging.
The best-evidence synthesis iteration approach has been developed in New Zealand (Alton-Lee 2004) and applied in a review of research on effective pedagogy in mathematics (Anthony & Walshaw 2007). This was prefigured by an earlier, more conventional review relating to science education.

Key features of the BESI protocol and process are an emphasis on inclusion, in the twin senses of engaging professional users and recognising student diversity; and an anticipation of implications, both through this engagement of users and by taking a broader systemwide perspective.
Compared to the EPPI reports (where a technical version provides a very detailed account of each stage in the review process) the BESI reports are more opaque in describing the operation of the review process and in accounting for the findings presented. A crucial part of the BESI process seems to be the identification of “seminal”, “landmark” studies which provide detailed accounts of linkages between pedagogy and outcomes “that pinpoint in unique ways how quality teaching might be characterised”.

Another feature of the BESI review is its rather ready rejection of a perhaps unnecessarily inflexible concept of “hard-and-fast rules about what methods and strategies work best” in favour of a more operationally ambiguous (and empirically untestable) “set of common, underlying pedagogical principles that appear to hold good across people and settings.”
Both EPPI and BESI, then, are relatively wide-ranging in approach: some of their distinctive characteristics are shown above.

Meta-analysis, by contrast, is a very specific approach which uses statistical methods to summarise findings about the effects of types of teaching process on types of learning product.

The following slides will sketch the three recent meta-analyses that have been conducted of effective pedagogy in mathematics and/or science.
Slavin and colleagues employ a variant of meta-analysis which (somewhat confusingly) is also termed “best-evidence synthesis” (although it is the NZ version which is the latecomer). They sought to identify effective teaching programs for mathematics at different school levels. The conceptual framework for the review is a conventional process-product one in which process is identified with type of teaching program, and product with student achievement outcomes.
Schroeder and colleagues sought to identify effective teaching strategies for science. The conceptual framework for the review is again a conventional process-product one in which process is identified with type of teaching strategy, and product with student achievement outcomes.

There are some parallels with the Slavin et al. process model, notably Collaborative Learning, Instructional Technology, and Direct Instruction (of which this review reports that no studies were found), but other quite substantial discontinuities.
The Seidel & Shavelson meta-analysis was part of a research effort that aimed to compare the efficacy of alternative pedagogical and psychological frameworks for conceptualising components of teaching and learning. It found that a process framework based on cognitive models of teaching and learning was relatively discriminating. Moreover, their product model was more wide-ranging, covering learning process and motivational-affective outcomes as well as cognitive outcomes. Serendipitously for the present analysis, Seidel & Shavelson reported separate findings for mathematics and science.

### Meta-analytic synthesis - components

**Publications**
- *Teaching effectiveness research in the past decade* (Seidel & Shavelson 2007)

**Conceptual framework**

**Process: Component types**
- Goal setting and orientation
- Execution of learning activities
  - Social interactions/direct experiences
  - Basic information processing
  - Domain-specific information processing
- Evaluation of learning
- Regulation/monitoring

**Product: Outcomes**
- Learning process
- Motivational-affective
- Cognitive

Disaggregates results to provide distinct mathematics and science findings
This table shows some important details of method for the three meta-analytic syntheses, and highlights certain important differences in these. Broadly, the Schroeder et al. review was relatively relaxed in its inclusion of studies in relation to study duration, research design, effect type, and outcome measure, while the Slavin et al reviews were relatively severe on study duration, research design and effect type, and somewhat conservative on outcome measure.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Study location</strong></td>
<td>Restricted to US</td>
<td>Unrestricted</td>
<td>Unrestricted but mainly US</td>
</tr>
<tr>
<td><strong>Study duration</strong></td>
<td>Unrestricted</td>
<td>Unrestricted</td>
<td>At least 12 weeks</td>
</tr>
<tr>
<td><strong>Research design(s) accepted</strong></td>
<td>Single group expt. or experimental comparison Prior control not required</td>
<td>Correlational survey or experimental comparison Prior control required</td>
<td>Randomised or matched experimental comparison Proximal prior profiles Prior control required</td>
</tr>
<tr>
<td><strong>Effect types</strong></td>
<td>Absolute &amp; Relative</td>
<td>Relative only</td>
<td>Relative only</td>
</tr>
<tr>
<td><strong>Outcome measures accepted</strong></td>
<td>Unrestricted: Generally researcher developed</td>
<td>Unrestricted: Standardised and researcher developed</td>
<td>Screened for intervention specificity: Mainly standardised</td>
</tr>
</tbody>
</table>
Surprising discontinuities between the reviews

- Differences in search strategies and screening criteria produced significant lack of overlap in the included studies
  - e.g. 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 latter
  - e.g. none of the 4 studies of the STAD program included by Slavin et al. which were eligible, by virtue of publication date, for the Seidel & Shavelson synthesis, featured in the latter
- Even where conceptual frameworks apparently coincided, crucial differences in the classification of studies emerged
  - e.g. IMPROVE classified as cooperative learning by Slavin et al., but as domain-specific information processing by Seidel & Shavelson

Some surprising discontinuities emerged between the reviews in their coverage and classification of studies.

Nevertheless, we will now focus on some key areas in which the findings, not just of the three meta-analyses, but of the other reviews, are at least commensurable and in some respects convergent.
I have used the term “domain-specific inquiry” to characterise a type of pedagogical construct that emerges (in the variants shown under “Teaching construct”) favourably across the reviews; and across both mathematics and science.

The (red-shaded) substantial effect sizes show consistent findings about positive impact on student achievement outcomes.

The (grey-shaded) evidence gap from the Seidel & Shavelson review suggests that very few process-product studies have looked at attitude outcomes in mathematics; at the same time, the best-evidence synthesis iteration by Anthony & Walshaw finds a favourable impact here. In science, the Seidel & Shavelson review found substantial evidence of positive impact on student attitude outcomes.
While the findings across reviews were commensurable on what I shall term “cooperative groupwork” they were not entirely convergent. In particular, there was a striking difference between the findings of the meta-analyses on whether cooperative groupwork is effective in relation to student achievement with the (red-shaded) effect sizes in the Slavin et al. and Schroeder et al. reviews suggesting positive impact and the (orange-shaded) effect sizes in the Seidel & Shavelson review suggesting null impact.

One reason for the difference may be that Slavin et al. restricted themselves to studies of interventions where teachers were in receipt of professional development, and the conclusions of Bennett et al. suggest that this is crucial. Another may be the restriction of the Seidel & Shavelson review to more recent studies (particularly in view of their preliminary finding that the measured effectiveness of pedagogical components may decline from one decade to the next). Equally, there is a note of caution in the finding from the best-evidence synthesis iteration by Anthony & Walshaw.

Finally, attitude outcomes appear to be much more favourable in science than in mathematics.
These findings have informed development of the *epiSTEMe* pedagogical model. In particular, they are reflected in the central role of domain-specific inquiry in the exploratory phase, and in the coupling of small-group and whole-class discussion (so that the latter can help model the patterns of interaction desired in the former, and structure the generation of findings from it).


Bennett, J., Lubben, F., Hogarth, S., & Campbell, B. (2004). A systematic review of the use of small-group discussions in science teaching with students aged 11-18, and their effects on students’ understanding in science or attitude to science. EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.


