IMA International Conference on Barriers and Enablers to Learning Maths:
Enhancing Learning and Teaching for All Learners
M.A. Hersh and M. Kotecha (eds.)

WHY 'WHAT WORKS' DOESN'T WORK IN PRACTICE, AND WHAT MIGHT WORK BETTER

Anne Watson

University of Oxford Department of Education
15 Norham Gardens, Oxford, OX2 6PY
Email: anne.watson@education.ox.ac.uk

Abstract: In this paper I critique a top down 'what works' approach to improving mathematics teaching and learning in schools, giving examples from a variety of sources; then I use as a case study my own experience from recent classroom teaching to suggest an alternative view of improvement.

Keywords: randomised control trials, 'what works' approaches, international comparisons, conceptual coherence, mathematics teacher knowledge.

1. Introduction: why a 'what works' approach is suggested

Current experience in English schools is heavily filtered through a marketplace approach to education, in which everything has to be measured and mechanisms have to be found to increase the measures assumed to be desirable. In a philosophy that everything of value can be measured, it is politically necessary to have a quantitative approach to controlling peoples' actions. This has led to the direction of public funds towards finding out 'what works' to improve students' grades in mathematics, as if there are globally reliable 'better' and 'worse' ways to teach. In its extreme form this requires randomised controlled trials with pre-and post test after field development.

Academic credibility is lent by Goldacre (2013) who claims that methods applicable to medicine and other sciences are also applicable in the complex world of education. To be clear, Goldacre's approach is that teachers can use randomised control trial techniques to find out for themselves 'what works' in their context, and that there are no 'one size fits all' findings in local situations. Goldacre himself does not suggest that global 'what works' findings have any purchase in complex situations involving human judgement and variability. However, policy-makers often use his ideas to justify the use of RCT methods to test ideas outside the original site of development.

It could be said that prescriptive education policy itself is a quantitative experiment, either a comparative study in which international test results are used as the measure, or a longitudinal study in which we expect a year on year improvements as prescriptions develop and change. When international comparisons began to show comparative weaknesses in achievement, it became clear that the traditional procedural skills-based approaches which had been used on hundreds of thousands of students over decades in many countries were an experiment that had not worked. In the First International Mathematics Study in the 60s, which was a traditional type of test, US achievement was worryingly low whereas England was 6th out of 12 and Scotland 7th. Results suggested that countries where 'new maths' had been introduced, as it had in England, did slightly better in traditional calculations than those who still used predominantly traditional methods. Twenty years later, the Second International Mathematics Study pointed to relative weaknesses of knowledge from different countries. It was possible to work out rankings and England and Scotland were around the middle, this time of 20 countries. This study gave educators an agenda to find out how to help more students understand the weakest topics, but instead the system was in the grip a national curriculum and testing regime; an opportunity for intelligence-driven change was lost (Brown, 1996).

Seen as a longitudinal study, efforts to improve mathematics teaching and learning are mixed. GCSE and A-level results in England in mathematics have increased consistently - measures of compliance of teaching with tests. During the last ten years, compliance of teachers with an imposed style of teaching has increased due to the effects of the inspection regime. At every key stage, students' test results rose year on year. While these success statistics have risen, our absolute performance in international tests has barely changed and may even be lower in secondary schools. Worse, a study of English students carried out by King's London, using the same test items and a similarly large sample as was used 30 years before, shows that there has been a significant lowering of achievement in the important areas of algebra and ratio, and a doubling to 15% of the proportion of students who, in KS3, can do very few questions even at a basic level (Hodgen et al. 2014).

Meanwhile a measurement culture has been promulgated through: national curricula and testing; legislation about practices; inspection and training regimes; teaching schemes and materials. All these assume that we know 'what works' and all that is necessary is to ensure that everybody carries it out, and those who do not are forced to change. For policy makers an ideal situation would be that the measures of compliance, namely test scores and performance indicators, rise; measures of non-compliance should fall. This way of evaluating innovation is not unusual at policy level - claims about the success of the Strategy to raise standards are mainly in terms of policy compliance (even called 'deliverology': Barber, Moffit and Kihn, 2010), not in terms of educational benefit in mathematics teaching and learning. Those in England who succeed in school mathematics at the higher level, through compliance, are nevertheless seen to be ill-prepared for university courses (Smith 2004).

Current policy aims at new improvement methods by, in Scotland, a grounded approach to curriculum reform deciding what is desirable and discussing with practitioners how to achieve it. In England the whole curriculum and assessment system is being recalibrated to follow the (untested) US example of government-defined core essential knowledge, and an (untested) accumulation of the curricula of countries more successful in international tests.

2. Critique of the 'what works' approach

The model of applying 'what works' assumes a homogeneous student body, and homogeneous teaching, so that roll out of successful small projects might also be successful at a larger scale. In practice, homogeneity is never achievable and in a diverse society may not even be desirable. As Ruthven et al. (2013) have shown rigorously, a successful innovation strategy can easily lose its effect when carefully applied to a larger group, when teachers who may not fully understand the underlying rationale take it on and use it in situations beyond those in which it was created, or when local adaptations are made to fit existing practices that are already valued and known to 'work'. By contrast, studies that focus on understanding what children can do in controlled circumstances, taught by a research team (e.g. Bryant and Nunes 2010) provide information about what is possible, and insight into difficulties in learning particular concepts, but cannot provide a recipe for effective teaching. Many 'what works' studies that do show success at small scale show no statistical significance at larger scale when rolled out beyond the innovation team (EEF 2015). Success is sometimes defined by tests that are valid internally to the aims of the innovation, and sometimes more broadly. In one recent study success was 'measured' by the capability of schools to conduct and act on their own evaluations, as Goldacre's approach would indicate (EEF 2015). Studies that do show statistical significance may also show small effect sizes (Hattie, 2013) because it is the fact of intervention and not its content that may have made a difference. The so-called Hawthorn effect is alive and well throughout education, because the enthusiasm of any innovators can always outweigh any dulling effect of non-optimal practice, such repetitive procedural work in mathematics. This could explain why, in the First International Mathematics Study in the 1960s it was found that, of the 12 countries that took part, those whose students did best in traditional arithmetic were those who had recently adopted the 'new maths' into their curricula including England, and hence had recent invigoration. Another problem is timescale. Education is for life, not merely for the next testing event. A recent RCT evaluation on Mathematics Mastery (MM) (EEF 2015) used a generic measure of mathematical achievement rather than a measure that was only valid for the project. Thus students in the innovation group were tested on topics that they had not been taught as well as those which were the focus of the programme. The outcomes showed that MM students did slightly better across the whole test than the control groups which had covered the full range of topics, but these differences were not statistically strong. However, the aim of MM is longer term improvement over several years, by focusing in depth on key ideas, so results over one year could never have given a fair result.

Innovations do not always show the same qualities when transferred to new contexts. A widely-promulgated teaching method developed in California called Complex Instruction was adapted and put into practice in a small trial in the UK with six participating schools and a control group of matched schools (Sebba et al. 2012). The method focused on the development of groupwork and the innovation was tested using gains in national mathematics test scores. Over the whole cohort there were no significant differences between the CI and the comparator groups of schools in terms of mathematics learning, and in individual matched comparisons the strongest significance was in a pair where the comparator school did better than the CI school. The evaluation report focuses not on the differences in mathematics learning but on qualities of the implementation of the innovation itself. This is fairly common among such studies, that the aim seems not to improve mathematics learning directly but to get more engagement, or more discussion, or a more positive attitude, or greater inclusion, with a belief that improvement in learning will follow. Another common assumption is that if weaknesses in implementation are addressed, learning would have been stronger according to the chosen measure. Typically, teachers involved in such projects grow in confidence across their teaching, even if they reject aspects of the project expectations. Currently, knowledge of the positive Hawthorn effects of *any* innovation leads to an agreement between teachers and policy-makers that a feature of successful teaching is that teachers work in teams to reflect on and develop their own practice, often using action research and/or collaborative methods to do so. By contrast, government imposition may not have this invigorating effect, and the architect of the Strategy has been reported as claiming it was a mistake to underestimate the negative effect on teachers (Wilby 2011).

Another problem with an RCT approach to 'what works' is that it only addresses two 'cells' of the practice/outcomes relationship (Figure 1). At rollout scale it is assumed that correct implementation should lead to success, and failure in implementation should prevent success; this is behind the recent inspection focus in England on lesson structure and teaching behaviour. Teachers who did not use expected structures had to justify their practice while those who used expected methods, possibly with less success, did not. RCT used as an evaluation method does not encourage researchers or practitioners to look at other ways of teaching that are successful and describe those in helpful detail, nor does it give enough attention to situations where full implementation does not lead to successful learning, and its outcomes depend on the measuring instrument. Instead, it encourages fluid interpretation of 'success' (e.g. "they asked more questions") rather than close identification of what really makes a difference in learning. Furthermore, the model does not encourage researchers to seek for other factors that may be common among both the more successful implementors and non-implementors that may have little to do with the project. Rather than segmenting the research question according to implementation and non-implementation, more may be learned from segmenting those teachers whose students learned more maths and those who did not and treat these as further conjectures to be tested in practice. For individual teachers, however, much can be learnt from *all* outcomes of a locally constructed RCT study.

	Implementation	No implementation
Successful learning		
Not successful learning		

Figure 1 A practice/outcomes relationship

It is worthwhile, therefore, to look for common features among situations that appear to promote strong mathematical learning. In an attempt to *explain* differences between students' learning in high achieving jurisdictions and the US, the TIMSS seven nation study (http://nces.ed.gov/pubs2003/2003013.pdf) found that there were very few observable and quantifiable features in common when looking at videos of typical lessons, most notably between Japan and Hong Kong who were the highest achievers at the time. This process has continued across many studies of many groups of countries and a research industry has developed so that we now know a lot more about the huge variety of mathematics education systems and practices, but not a lot more about any common strengths in higher achieving countries. The researchers therefore convened a 'mathematics quality' panel to comment on the mathematical content of lessons as possibly the critical feature in successful teaching. This, I claim, is the most important but least developed outcome of the TIMSS seven-nation study. They found that, whatever the observable teaching characteristics, high achieving countries shared a strong sense of coherent conceptual development in their teaching, and taught harder maths, and paid more attention to conceptual work rather than procedural, and gave a rationale for what was being taught, included generalisations as well as examples, and made clear connections and continuations in their presentations and plans. It is remarkable that what dominates some areas of research literature is the implementation of so-called 'reform' methods of teaching rather than the enactment of coherent conceptual development found to be important in this study. I conjecture that the difficulties in doing this are due to disconnections between psychologists who map students' micro-elements of understanding in non-teaching contexts, teachers who work in the generic contexts of schools rather than within a subject specific culture, and designers who report how a particular concept might be taught. These disconnections are major obstacles and are exacerbated by the pressure on academics to publish, so that early career researchers, or even early career mathematicians, publish short term outcomes of their innovations or teaching methods rather than waiting to see the longer term effects. Another problem is that teachers who have developed effective ways to teach and want to report them rigorously for others have to negotiate the world of social science research, and even RCT, when their initial training was as mathematicians.

The 'what works' philosophy positions individual teachers and students as the problem, and systems and policies as the solution. In what follows I do the opposite; I position systems and policies as the barriers and an individual teacher (myself) and her students as the enablers.

3. Working in a 'what works' environment

After 13 years as a secondary school mathematics teacher, and 19 years academic work on mathematics education, I recently returned as a volunteer to teach in a comprehensive school that had high levels of mathematical achievement for most, but not all. I took main charge of teaching a year 7 class of children who were not achieving national expected levels in mathematics. The school had several practices in place that have all, at some time or other, been presented to teachers as 'what works' policies justified by generic school improvement advice:

- 'one system fits all' behaviour management systems;
- schemes of work that match national guidelines;
- regular testing against national standards;
- class sets of several textbooks and DVDs;
- additional teaching for those who are not achieving national standards.

Good research about positive effects of whole school systems of rewards and consequences is hard to find and results vary between systems, but it appears that intervention is more effective when it is designed for particular students (EEF 2014). Evidence for its success in the UK often comes from the companies who sell relevant software, and is based on reports from pleased users. In the US, evidence from external researchers suggests that it sometimes leads to more peaceful schools (where the starting point was sometimes extreme violence and knife and gun crime) but those who end up being excluded from school are likely to be from disadvantaged backgrounds. Most of my students were from disadvantaged backgrounds and some reached a 'consequences' level of exclusion from lessons with the first few weeks of school, which seemed to me to be unhelpful.

Research about textbooks, tests and other materials that shape the curriculum in UK shows consistent weaknesses (e.g. Haggarty & Pepin 2002). They present mathematics as fragmented, test-driven, and without the coherent conceptual development evident in textbooks from some more successful countries. It is also easy to find examples of mathematical errors, or misleading statements, in many widely-available resources, even from well-established publishers. For example the resources available to me were particularly limited, and sometimes misleading, in their treatment of multiplicative reasoning, area and linear functions.

Research about the success of additional teaching mainly points to one-to-one input supported by diagnostic testing, rather than extra classes or a satellite curriculum provided on the basis of SATs results (e.g. Holmes and Dowker 2013). Most of the research focuses on primary students but given the weaknesses some of my students showed in number sense the same is likely to apply. Half of my students had extra mathematics lessons in a satellite group with a non-specialist teacher, selected for this according to their low English SATs scores, with no diagnostic testing for mathematics. The curriculum for these extra lessons was mainly whole number calculation practice, multiplication tables, and consolidation games appropriate for the primary curriculum.

As their teacher, I was expected to work within these practices. I was no longer employed as an academic, so there was no pressure to publish; instead I wished to apply my interpretation of ideas about learning, informed by Geary's observation that ' … roughly half of the children who had been identified as having a learning problem in mathematics did not show any form of cognitive deficit …' (Geary 1994 p.157). This supported the view that many children who had been consigned to 'bottom sets' for mathematics were there because of problems in teaching, the educational system and continuity of schooling rather than cognitive problems (see also Denvir, Stolz, and Brown, 1982).

3.1 Teaching towards proportionality

For all students in most countries, the early years of secondary school are a key time for learning about proportionality. The school scheme of work allowed about five or six lessons for ratio and proportion compared to the 22 lessons spread in two chunks over two years I knew were typical in Japan, for example. There was also a test due after six weeks of school, which would address the full spread of national curriculum mathematics. Since all the students had different prior knowledge, being from different primary schools and different prior levels of attainment (the 'bottom set' being necessarily a wider distribution of attainment than other groups) teaching step-by-step was not an option; similarly, individual programmed learning was inappropriate due to low reading levels, low levels of self-management, and the need to develop ways to talk about mathematical ideas through working on the same problems. More could possibly be achieved by addressing the mathematical concept of multiplicative reasoning as if they had no cognitive deficits than by assuming deficits and therefore teaching simpler concepts. I therefore decided to approach proportional reasoning by first ensuring they all had experience of multiplication as scaling, comparing fractions of quantities, and having appropriate notations for these ideas. I therefore had to design tasks that coordinated several different aspects of prior knowledge into coherent learning experiences for the students, while simultaneously providing key learning points for individuals that might help them move on in knowledge.

My focus was on what was available to be learned, and how it can be made available through various representations, language forms, physical materials, visual images and experiences that require transformation and interaction between these. This approach is embedded in the content and design of prescribed textbooks in China and Finland, and research being undertaken in Sweden and elsewhere (Koichu, 2013; Runesson et al., 2006;Sun, 2011). This work is showing signs of success in South Africa, where teacher knowledge can be very weak (Venkat & Naidoo 2012), and is in accord with approaches developed by Gattegno (e.g. 1970, see also Hewitt (1994)). Its key feature is that the mathematical tasks offered to learners should focus them on the essential structures of mathematics, and the task of lesson design is to work out how most students can be given experience of these (Watson & Mason, 2006).

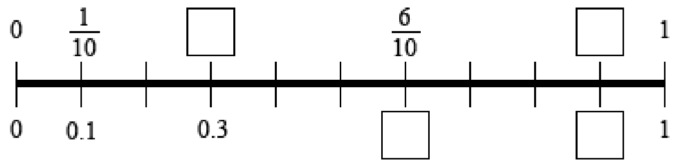
I can only give a flavour of it here, and I am not presenting this is perfect, or 'the best' way to teach. I do know, however, that it 'worked' even within the difficult school environment of whole school generic structures described above in the sense that all students made progress.

Number of equal pieces #	Fraction of a metre	Measurement in centimetres	Percentage of a metre	1 ÷ #	Decimal fraction
2					
4					
8					
5					
10					

Figure 2 A format for relating metre-long paper strips to measurements.

3.1.1.Task A: Metre-long strips of paper were to be folded into two, three, four etc. equal pieces and measured using a metre stick. Results were to be entered on a record sheet (Figure 2) and compared. The aims were to: develop a length image for fractions; associate fractions with division; link to percentages and place value for decimals; draw out existing knowledge; expose typical confusions about relative sizes of unit fractions and decimal equivalents. I used the strategy of providing formats that revealed conceptual connections and patterns as a regular device for helping students shift between representations to help conceptual understanding (Duval 2006).

1. Fill in the missing labels. [Extension: extend the line to the right and put some more labels of your own]



2. Add these pairs of tenths and look for those that give the same answer. Why do they give the same answer? [Extension 1: turn them into decimals using the numberline Extension 2: make up some 'adding tenths' that give an answer bigger than 1]

$\frac{2}{10} + \frac{4}{10} =$ $\frac{9}{10} + \frac{1}{10} =$ $\frac{8}{10} - \frac{1}{10} =$

$\frac{3}{10} + \frac{3}{10} =$ $\frac{2}{10} + \frac{8}{10} =$ $\frac{8}{10} - \frac{5}{10} =$

$\frac{1}{10} + \frac{5}{10} =$ $\frac{7}{10} + \frac{3}{10} =$ $\frac{8}{10} - \frac{2}{10} =$

$\frac{4}{10} + \frac{2}{10} =$ $\frac{4}{10} + \frac{6}{10} =$ $\frac{8}{10} - \frac{8}{10} =$

Figure 3 A collection of examples to work on involving tenths

3.1.2 Task B: From observations of the activity prompted by task A, further tasks were designed to emphasise particular conceptual aspects. For example, a later lesson involved a physical number line (string with pegs for points) with whole class participation in chanting counting along it in tenths in both directions, placing multiples of one tenth. This was followed by a personal task in class (Figure 3) with the specific verbal instruction to think about the numberline and remember the chanted counting, using multi-sensory episodic memory to make the conceptual connections. Another feature is that it is worthwhile reflecting on the collection of questions which has constraints, repetitions, and related variations between the examples (Watson and Mason 2006). The principle of controlling variation is at the heart of teaching in China and some other countries (Sun, 2011).

Although I chose to focus mainly on tenths when treating fractions as numbers because seeing them as lengths on the numberline supported their understanding of place value, I was also concerned that they should connect fractions with division of quantities. The curriculum traditionally puts addition and subtraction *of* fractions before multiplication *by* fractions, but with this class I decided to work mainly on multiplicative reasoning because if its uses across and outside mathematics.

A, E, L, H, J and K had five rectangular chocolate cakes to share equally between them. They did it in a really clever way. Here it is described as division. The answers are fractions:
3 ÷ 6 = ... of a cake each 2 ÷ 6 = ... of a cake each
Draw diagrams to show what they did.

So each person gets: ─ + ─ = of a cake each

Figure 4 Re-enacting 'sharing out a quantity' using fraction format

3.1.3 Task C: Students were given several quantities of food to share between different numbers of students. They presented their solutions. I rewrote their methods overnight in a format for them to create on a personal record of the lesson, using fractions notation where possible. This move was to harness episodic memory and introduce fraction notation for formal recording. Figure 4 shows an example of such a 'fractions of quantity' situation which had first been acted out in the classroom.

3.1.4 Task D: The move to thinking of multiplication as scaling was made through a task in which students had to decide the size of a giant from some articles of clothing and personal possessions, and from this design further objects and meals for it. This is a well-known task (e.g. Streefland 1984) and I provided a format to draw attention to the constant of proportionality (Figure 5).

	Giant		Human			Human				Giant
Shoelace		÷		=			×		=	
Tie		÷		=			×		=	
Bus pass length		÷		=			×		=	
Bus pass width		÷		=			×		=	
Stamp side		÷		=			×		=	
Stamp width		÷		=			×		=	
Scarf		÷		=			×		=	
Sock		÷		=			×		=	

Figure 5 Formatting to demonstrate the constant of proportionality

3.2 Learning outcomes

Mid-term the problem of testing arose: students had to take a test that was neither summative of what they had learned nor formative about what they were learning. I prepared by drawing on dual-process theory and giving them practice in providing reflective rather than intuitive answers (Leron & Hazzan 2009). I also allowed them to opt to take a higher level test when they finished the one that was assigned by 'the system'. In the event they all did much better in the test than they had five months earlier on similar tests, with a long holiday and a change of school in between. I therefore claim that my teaching was effective according to authentically valid assessment. There were several unexpectedly high scores. I focused on processing because I was confident about their underlying reasoning capabilities and did not want to disrupt the conceptual flow with revision sessions. As with all the teaching described above, my belief was that the combined minds of myself and the students could 'beat the system' and show progress even without curriculum coverage, or test practice.

4. Barriers and enablers

My approach has been to position systems and management decisions as potential barriers, and the actions of an individual teacher and group of students as enabling. Rather than seeing my work as the effects of having an experienced teacher, I hope it will be seen as an example of research-informed practice in action, even within a less than ideal organisational situation. I drew explicitly on variation theory, dual-process theory, international comparisons of conceptual coherence of lessons, research about the centrality of multiplicative reasoning, and research about cognitive deficits. I used task design knowledge and research about the role of exemplification in conceptual learning. There has been a recent shift in mathematics education research towards the nature and importance of teachers' mathematical pedagogical knowledge (Rowland & Ruthven 2010; Watson 2008) which would include the kinds of knowledge in which I drew. Simultaneously there has been a rapid shift in England away from university involvement in teacher training and development, yet few of the theoretical sources I used to design my teaching are available easily to teachers and schools. It will be interesting to see the effects of these contradictory shifts.

References

Brown, M. (1996). FIMS and SIMS: the first two IEA International Mathematics Surveys. *Assessment in Education*, *3*(2), 193-212.

Bryant, P. & Nunes, T. (2010). *Children's understanding of probability: an intervention study.* nuffieldfoundation.org/childrens-understanding-probability-intervention-study.

Denvir, B., Stolz, C., & Brown, M. (1982). *Low attainers in Mathematics 5–16: problems and practices in school.* London, Methuen.

Duval, R. (2006). A cognitive analysis of problems of comprehension in a learning of mathematics. *Educational studies in mathematics*, *61*(1-2), 103-131.

EEF (Education Endowment Foundation) (2014) *Behaviour intervention toolkit.* educationendowmentfoundation.org.uk/uploads/pdf/Behaviour_Intervention_Toolkit_ references.pdf

EEF (2015) *Mathematics Mastery.* http://educationendowmentfoundation.org.uk/projects.

Gattegno, C. (1970). *What we owe children: The subordination of teaching to learning*. Educational Solutions.

Geary, D. C. (1994). *Children's mathematical development: Research and practical applications*. American Psychological Association.

Goldacre, B. (2013) *Building evidence into education.* media.education.gov.uk/assets/files/pdf/ b/ben%20goldacre%20paper.pdf)

Haggarty, L., & Pepin, B. (2002). An investigation of mathematics textbooks and their use in English, French and German classrooms: Who gets an opportunity to learn what? *British Educational Research Journal*, *28*(4), 567-590.

Hattie, J. (2013). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Routledge.

Hewitt*, D*. (1994) *The principle of economy in the learning and teaching of mathematics,* unpublished Ph.*D*. thesis*,* Milton Keynes, Open University.

Hodgen, J., Küchemann, D., Brown, M., & Coe, R. (2009). Children's understandings of algebra 30 years on. *Research in Mathematics Education*, *11*(2), 193-194.

Holmes, W., & Dowker, A. (2013). Catch up numeracy: a targeted intervention for children who are low-attaining in mathematics. *Research in Mathematics Education*, *15*(3), 249-265.

Koichu, B. (2013). *Variation theory as a research tool for identifying learning in the design of tasks*. Plenary panel at the ICMI Study-22 Conference, The University of Oxford.

Leron, U., & Hazzan, O. (2009). Intuitive vs analytical thinking: four perspectives. *Educational Studies in Mathematics*, *71*(3), 263-278.

Rowland, T., & Ruthven, K. (2011). *Mathematical knowledge in teaching* (Vol. 50). Dordrecht, Heidelberg, London, New York: Springer.

Runesson, U. (2006). What is it possible to learn? On variation as a necessary condition for learning. *Scandinavian Journal of Educational Research*, *50*(4), 397-410.

Ruthven et al. (2013) *EpiSTEMe Final report*. www.educ.cam.ac.uk/research/projects/episteme/ epiSTEMeFinalReport.pdf

Sebba, J., Altendorff L., Kent, P., Boaler, J. (2012) *Raising expectations and achievement levels for all mathematics students: final report.* http://esmeefairbairn.org.uk/news-and-learning/publications.

Smith, A. (2004). *Making mathematics count: The report of Professor Adrian Smith's inquiry into post-14 mathematics education*. London, HMSO

Streefland, L. (1984). Search for the roots of ratio: Some thoughts on the long term learning process (Towards... a theory). *Educational Studies in Mathematics*, *15*(4), 327-348.

Sun, X. (2011). "Variation problems" and their roles in the topic of fraction division in Chinese mathematics textbook examples. *Educational Studies in Mathematics*, *76*(1), 65-85.

Venkat, H., & Naidoo, D. (2012). Analyzing coherence for conceptual learning in a Grade 2 numeracy lesson. *Education as Change*, 16(1), 21-33.

Watson, A. (2008). School mathematics as a special kind of mathematics. *For the Learning of Mathematics*, 3-7.

Watson, A., & Mason, J. (2006). Seeing an exercise as a single mathematical object: Using variation to structure sense-making. *Mathematical Thinking and Learning*, 8(2), 91-111.

Wilby, P. (2011) *Mad professor goes global.* www.theguardian.com/education/2011/jun/14/michael-barber-education-guru.