Using longitudinal, cross-system and between-subject analysis of the TIMSS study series to calibrate the performance of lower-secondary mathematics education in England

Kenneth Ruthven

University of Cambridge

Evidence from the TIMSS study series is used to calibrate trends in lower-secondary mathematics in England between 1999 and 2007, relative to other educational systems, and in comparison with science. Over this period the proportion of students displaying high achievement in mathematics rose, apparently in response to reforms associated with the national *Strategy*. However, the proportion of students displaying positive attitude to mathematics fell substantially. In both respects, performance in mathematics continued to compare unfavourably with that in science.

Performance trends; international comparisons; mathematics education; science education; student achievement; student attitude.

Introduction

The focus of this paper is on using findings from the TIMSS international study series to examine trends in the performance of the English educational system within a framework of cross-system and between-subject comparison. This work has been undertaken in the context of a current research initiative (Economic and Social Research Council [ESRC] 2006) intended to inform ongoing efforts to secure significant enhancements in young people's school achievement in science and mathematics, and significant increases in their participation in further study and employment in these areas. As part of the initiative, the *Effecting Principled Improvement in STEM Education (epiSTEMe)* project (Ruthven et al. 2010) is undertaking research-based pedagogical development, designed to be suited to implementation at scale within the English educational system, and aimed at improving student engagement and learning in early secondary-school physical science and mathematics.

In recent years, schooling in England has been reshaped by two waves of reform aimed at standardising educational provision and improving educational outcomes. From 1989, a national curriculum was introduced, and a system of national assessment at 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 performance, particularly in student achievement, and by launching a national programme of school improvement, supported by extensive professional development. From 1999, in particular, the teaching of mathematics in primary schools in England was expected to follow the approach to lesson planning, teaching method, and classroom assessment established by the national *Strategy* (Brown et al. 1998; Reynolds and Muijs 1999). Initially developed for use at primary level, this pedagogical model was extended to lower-secondary level both in mathematics (from the student cohort entering in 2001) and in science (from the cohort entering in 2002).

Reporting on implementation of the *Strategy* at lower-secondary level, the main trends noted by school inspectors in mathematics were towards "improvements in the planning of teaching, with a greater focus on learning objectives, the structure of lessons and teachers' use of questioning" (Office for Standards in Education [OfStEd] 2004, 21), and the introduction of "systems… for regular monitoring of pupils' performance, with action taken to help them improve" (OfStEd 2004, 24). The approach developed in science was similar: inspectors found that the recommended lesson structure was near universal, as was emphasis on teaching to explicit objectives (OfStEd 2004, 31), and that teacher assessment was "being increasingly used more successfully to guide pupils on how to improve their work" (OfStEd 2004, 29).

Using the TIMSS study series to analyse changes in system performance

International study series such as the Trends in International Mathematics and Science Study [TIMSS] provide a framework for examining educational trends in terms of cross-system and between-subject comparison. In addition, their estimates of student achievement are less open to inflation as a result of the "teaching to the test" that high-stakes national testing encourages. Furthermore, 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 further study (Ruddock et al. 2006); important for the particular concerns of the ESRC initiative with raising student participation in more advanced mathematical and scientific courses.

Equally, the TIMSS series usually allows an age cohort to be tracked from the elementary/primary level in one study to the middle/lower-secondary level in the next (as shown in Table 1). Thus the entire schooling of what has been termed the "earlier" English TIMSS cohort took place following the initial reforms but well prior to the subsequent ones associated with the *Strategy*. TIMSS provides less information about the "intermediate" cohort[1] which was the last to have virtually no experience under the subsequent reforms, while the "later" cohort was entirely schooled under these subsequent reforms. Thus any improvement in system outcomes as a result of the later 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 study series

Cohort	Entered primary phase	Studied at Grade 4 (Year 5)	Entered secondary phase	Studied at Grade 8 (Year 9)
Earlier cohort	1990	1995	1996	1999
Intermediate cohort	1994	*not studied*	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 and Foy 2008a; 2008b; Sturman et al. 2008). The findings for England will be situated within the distribution of results across

[1] The intermediate cohort was not studied at Grade 4 level in 1999, and did not complete the main attitude measures at Grade 8 level in 2003. There were also sampling weaknesses: England nearly satisfied guidelines for sample participation rates only after replacement schools were included. In addition, 3 of the other systems used as comparators here did not participate in the 2003 TIMSS study.

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 level[1].

Establishing benchmarks for student achievement and attitude

TIMSS defines (in terms of test scores) several "benchmarks" for student achievement. Here, the focus will be on what TIMSS terms the "high international" benchmark. In Mathematics at Grade 8 level, this is characterised in terms of students being able to "apply their understanding and knowledge in a variety of relatively complex situations", rather than, at the next benchmark, to simply "apply basic mathematical knowledge in straightforward situations". This provides, then, a suitable marker for the level of capability required for more advanced study in STEM fields. Likewise, in Science, students who achieve the high benchmark are characterised as being able to "demonstrate conceptual understanding of some [science]", compared to the lower "recognise and communicate basic scientific knowledge". Here then, the operational index of "[subject] achievement" in an educational system will be the percentage of students reaching this high achievement benchmark in the subject.

Turning from achievement in a subject to attitude towards it, the most relevant outcome measure available in the TIMSS study series is of "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 participation in more advanced study in STEM fields. Thus here, the operational index of "[subject] attitude" will be the percentage of students in an educational system reaching this positive attitude benchmark in the 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 students reaching the high achievement benchmark for science was 39% for the median system in 1999 and 38% in 2007. The median system change in this percentage between cohorts was +1. While the main analysis will focus on these cohorts because fuller information is available on them, the intermediate cohort will later help to triangulate and sharpen findings on achievement.

Examining between-cohort changes in system performance on student achievement

In science, the markers in Table 2 show little overall change in levels of achievement between the earlier and later cohorts. For England specifically, performance did not change significantly in absolute terms, and remained well above the median position and a little below the upper quartile in relative terms. The situation as regards mathematics is rather different. Overall levels of achievement drifted downwards

[1] 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 in relation to the United States, the portion is so small, and state-level education policies sufficiently distinctive, that both entities can reasonably be included as independent 'systems' within the analysis. Likewise, the relative independence of the Canadian provinces in educational matters justifies their treatment as distinct 'systems'.

between the cohorts. However, there was a marked improvement (the second largest such improvement) in the absolute performance of England. In relative terms, too, 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, whereas England's performance in science changed negligibly between the two cohorts, performance in mathematics improved, although not to a level matching science.

Table 2: Distribution of system performance on Grade 8 achievement: Proportion of students reaching the TIMSS high achievement benchmark[1]

Distribution	Mathematics			Science		
	Earlier cohort scores	Later cohort scores	Between cohort changes	Earlier cohort scores	Later cohort scores	Between cohort changes
Upper quartile	65%	59%	+ 1	47%	53%	+ 3
Median	33%	32%	- 2	39%	38%	+ 1
Lower quartile	20%	17%	- 6	23%	22%	- 2
England	**25%**	**35%**	**+ 10 ↑**	**45%**	**48%**	**+ 3 $_{NS}$**

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 the 18 systems under consideration. Of these 8 systems, England had the largest improvement in mathematics achievement at both levels. Nevertheless, the improvement at Grade 8 level (of 10 percentage points) was notably smaller than at Grade 4 level (where the rise was 19 percentage points, from 24% to 43%). However, smaller but still substantial improvements at Grade 4 level in Hong Kong and Ontario did not feed through to Grade 8 level. These patterns suggest, then, that there is no straightforward relationship between changes at the two levels. Nevertheless, given the magnitude of the English improvement at Grade 4 level, it is plausible to conjecture that this contributed to the subsequent improvement at Grade 8 level.

Triangulating between-cohort changes in performance on student achievement

To investigate further, it is necessary to take account of evidence about the intermediate cohort, and helpful to triangulate TIMSS findings against the results of national testing (DCSF 2008). Over the period under review, English schools and teachers were under enormous pressure to improve the performance of students in national tests: for lower-secondary (KS3) assessment at the end of Year 9, level 6 represented the key benchmark of high achievement. The graphs for science and mathematics (Figure 1) show that the relative demands of the national and TIMSS benchmarks differ between subjects: in science, the level 6 benchmark in national testing is more demanding than the TIMSS high achievement benchmark; in

[1] Significance of differences between cohorts are those reported in the TIMSS studies, and indicated as follows: later cohort: ↑ significantly higher; $_{NS}$ not significantly different; ↓ significantly lower

mathematics, this pattern is reversed. In TIMSS terms, then, the expectations of English national testing are higher in science than in mathematics.

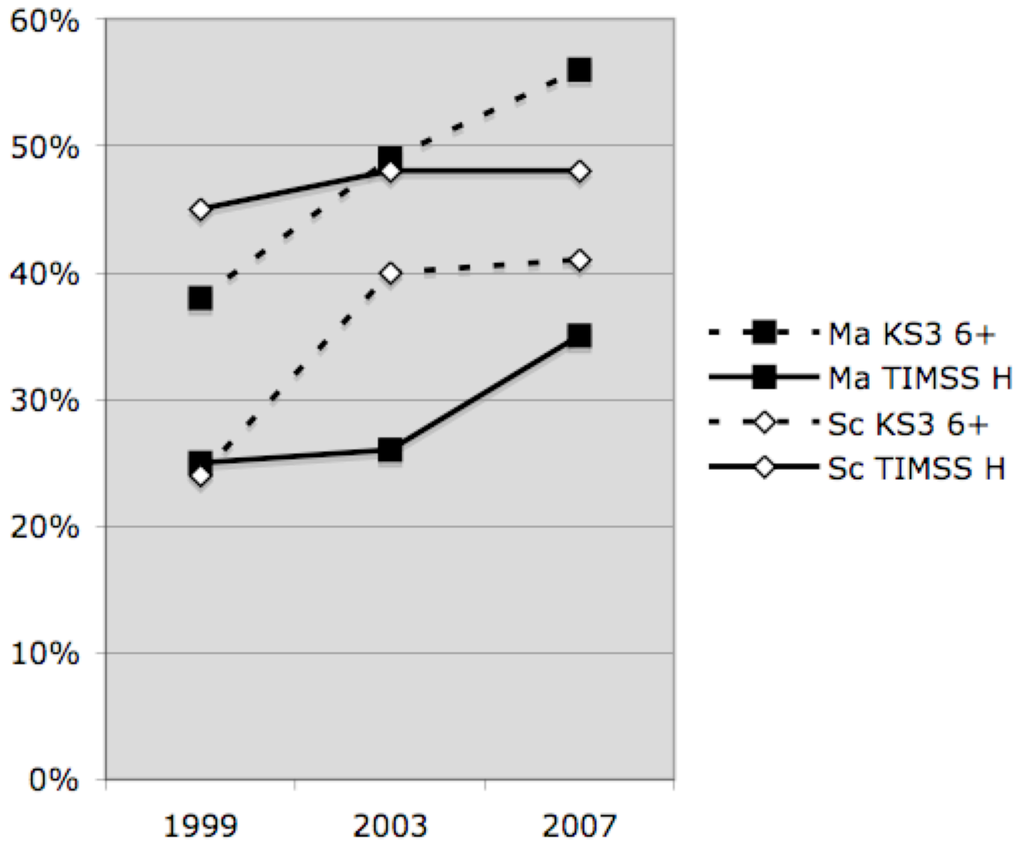

Figure 1: Change in English system performance between cohorts: Proportion of students achieving the TIMSS high benchmark compared to level 6 in KS3 tests.

On trends over time, the graphs show that performance on national tests improved markedly between 1999 and 2003 in both science and mathematics, while performance on TIMSS remained static: this suggests a fairly superficial form of improvement, 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 static in science, whereas it rose markedly in mathematics. This suggests that the reforms associated with the *Strategy,* which affected only the later cohort, led to a more fundamental form of improvement in achievement in mathematics, but not in science[1].

Examining between-cohort changes in system performance on student attitude

Turning to assessment of student attitude within TIMSS (Table 3), in both subjects, the markers for the overall distributions of system scores, and for the distribution of individual system changes, indicate that the predominant pattern was one of decline in performance. In both subjects England saw a very substantial fall between the two

---

[1] Some caution needs to be exercised over TIMSS 2003 results because of the sampling problems referred to in Note 2. Expert review of TIMSS items has also found their likely familiarity to English students to be rather greater in mathematics than in science (Ruddock et al. 2006). 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.

cohorts, markedly greater than the general trend. In science, 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). In mathematics, too, England displayed a 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, England performs more strongly in science than in mathematics, certainly in absolute terms, rather less clearly so in relative terms.

Table 3: Distribution of system performance on Grade 8 attitude: Proportion of students reaching the TIMSS positive attitude benchmark

Distribution	Mathematics			Science		
	Earlier cohort scores	Later cohort scores	Between cohort changes	Earlier cohort scores	Later cohort scores	Between cohort changes
Upper quartile	67%	59%	- 1	77%	68%	+ 1
Median	58%	47%	- 7	63%	55%	- 6
Lower quartile	46%	39%	- 12	59%	52%	- 10
England	**65%**	**40%**	**- 25 ↓**	**76%**	**55%**	**- 21 ↓**

Examining between-cohort changes in system performance on combined outcomes

In some systems, there was very little change in combined performance on achievement and attitude between cohorts, either in science (Figure 2) or in mathematics (Figure 3): in Tunisia (TN) for example (at the top left of both figures). 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, trend lines relating attitude to achievement across the systems as a whole indicate that higher achievement tends to be associated with lower 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 again show that this is predominantly due to changes (downwards) in attitude rather than (backwards) in achievement.

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%), but this is in line with the median decline across systems in each of these subjects. On achievement, England (EN) was the other strong riser in mathematics, but not in science. On attitude, as already noted, England fell very substantially in both subjects, at or near the extremes of decline.

The Massachusetts example shows that such a fall cannot be attributed to any 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. 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. However, whereas the pedagogical model promoted by the *Strategy* in England was shaped by older research on the effective teaching of basic skills (Reynolds and Muijs 1999), the *Standards* influencing reform in Massachusetts reflected more recent research on developing higher-order thinking (Massachusetts Department of Education 1999).

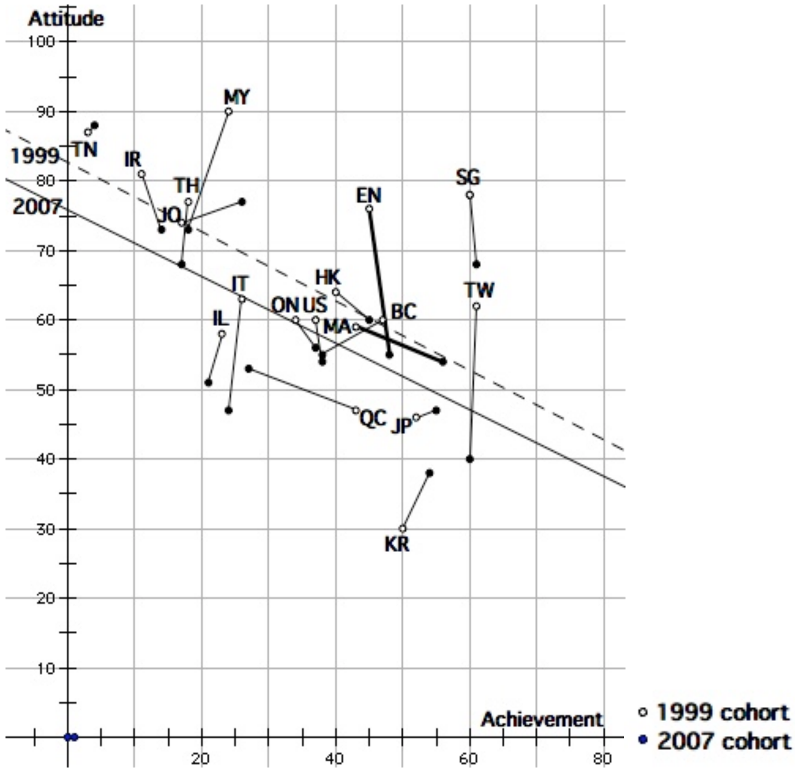

Figure 2: Change in system performance in Grade 8 science between TIMSS cohorts.

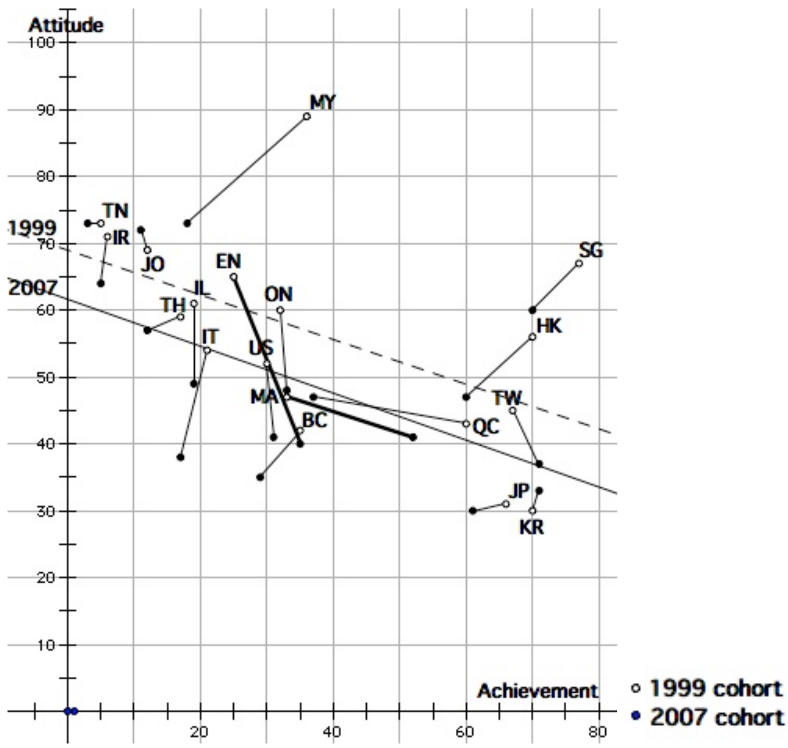

Figure 3: Change in system performance in Grade 8 mathematics between TIMSS cohorts.

Conclusion

This paper has shown how the TIMSS international study series can be used to construct a framework for cross-system and between-subject comparison capable of illuminating trends in system performance on student achievement and attitude. Applied to the case of English lower-secondary schooling, this challenges the optimistic picture from national assessment. Compared to other systems in terms of the TIMSS high achievement benchmark, fundamental gains by English students have taken place only in mathematics, and then only in response to the later reforms associated with the national *Strategy*. Compared to other systems in terms of the TIMSS positive attitude benchmark, declines in both mathematics and science have been exceptionally severe amongst English students. Finally, in these international terms, English performance remains lower in mathematics than in science.

These findings emphasise the importance of developing a better understanding of student affect and identity, a salient feature of many of the projects in the ESRC research initiative. Equally, new insights are clearly required into ways of raising student achievement: in England, there is scope for further improvement in mathematics from a still relatively average performance, and for improvement in science beyond a relatively high, but static, performance. Major recent shifts in English education policy – notably, revision of the national curriculum to reduce prescriptiveness; abolition of compulsory national testing at lower-secondary level; abandonment of centrally-driven school improvement – may now have created the conditions for projects such as *epiSTEMe* to collaborate with schools and teachers in undertaking pedagogical development towards these goals in the light of the most recent research on effective teaching and learning.

References

Brown, M., M. Askew, D. Baker, H. Denvir, and A. Millett. 1998. Is the National Numeracy Strategy research-based? *British Journal of Educational Studies* 46(4): 362-385.

Department for Children, Schools and Families [DCSF]. 2008. National Curriculum Assessments at Key Stage 3 in England, 2006/07 (Revised). *Statistical First Release* 06/2008. London: DCSF.

Economic and Social Research Council [ESRC]. 2006. *Targeted Research Initiative on Science and Mathematics Education*. Swindon: ESRC.

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

Mullis, I., M. Martin, and P. Foy. 2008a. *TIMSS 2007 International Mathematics Report*. Boston: TIMSS International Study Center.

——— 2008b. *TIMSS 2007 International Science Report*. Boston: TIMSS International Study Center.

Office for Standards in Education [OfStEd]. 2004. *The Key Stage 3 Strategy: evaluation of the third year*. London: OfStEd.

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

Ruddock, G., T. Clausen-May, C. Purple, and R. Ager. 2006. *Validation Study of the PISA 2000, PISA 2003 and TIMSS 2003 International Studies of Pupil Attainment* (DfES Research Report 772). London: DfES.

Ruthven, K., C. Howe, N. Mercer, K. Taber, S. Luthman, R. Hofmann, and F. Riga. 2010. Effecting Principled Improvement in STEM Education. *Proceedings of the British Society for Research into Learning Mathematics*.

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