A Comparison of PISA and TIMSS against England's National Curriculum

Burdett, N. and Sturman, L.

Contact: n.burdett@nfer.ac.uk
National Foundation for Educational Research,
The Mere, Upton Park, Slough, Berkshire SL1 2DQ
www.nfer.ac.uk

Keywords: International surveys; TIMSS; PISA; international comparisons

Abstract

This study updates an earlier NFER study validating the PISA and TIMSS test items in terms of their familiarity for secondary school students in England (Ruddock *et al.*, 2006). The study explored whether item familiarity (or unfamiliarity) might help to explain apparent differences in attainment between the PISA and TIMSS surveys in 2000 and 2003. This related study addresses the fact that differential performance in England seems to remain. While there are several potential reasons for this, the difference in familiarity and reading demand between PISA and TIMSS may still impact on performance. In addition, the national teaching and assessment contexts have changed since 2006.

This study looks at item familiarity of maths and science items in TIMSS 2007, PISA 2009 and PISA 2012 and in addition looks at the curriculum context, which was not covered in the previous study.

This study found that:

- The assessment frameworks for PISA and TIMSS cover the same basic content and skills as the national curriculum for mathematics and science in England.
- The PISA test items compare with the TIMSS test items and both compare with those familiar to students in England.
- The implication of this research is that PISA and TIMSS outcomes can be validly used for the study and comparison of the English educational system.
- Changes in performance of students in England over time are unlikely to be due to a change in the items or frameworks of either international study, although reading demand might play a small part.
- Higher apparent performance in TIMSS compared to PISA is unlikely to be caused by familiarity with the content areas or with the items included in the assessments.

Introduction

The original Validation Study published in 2006 looked specifically at the extent of familiarity of the TIMSS and PISA test items for students at the relevant ages to assess the extent to which the content might have affected outcomes. The study found that, while the items in both surveys were familiar to students, they differed in terms of the reading demand on students. Since then, further test items have been developed to replace those released after each survey cycle and attempts are being made to lower the reading demands in PISA. While there could be several potential reasons the difference in familiarity and reading demand

between PISA and TIMSS may still impact on performance. In addition, the national teaching and assessment contexts have changed since the 2003 cycles.

This study explores potential reasons for differential performance in PISA and TIMSS. These includes the extent to which test items are familiar to students, and other questions raised but not fully addressed during the original study, such as reading demand. It also explores the extent of curriculum match between the assessment frameworks for PISA and TIMSS, in comparison with the national mathematics and science curricula in England.

Method

As in the original study, a team of expert raters was recruited (two raters for mathematics and two for science). They carried out a curriculum matching exercise and a test item rating exercise.

The curriculum matching element was designed to ensure that like was being matched with like and therefore comprised seven comparisons (to account for the structure and timing of the PISA and TIMSS surveys, and for changes in the national curriculum over time). These were as follows.

For mathematics:

- The 1999 (revised 2004) key stage 3 (KS3) mathematics curriculum with the TIMSS-2007 Assessment Framework
- The 2007 KS3 curriculum with the TIMSS-2011 Assessment Framework
- The 1999 key stage 4 (KS4) mathematics curriculum with the PISA 2003 Assessment Framework.
- The 2007 KS4 curriculum with the PISA 2012 Assessment Frameworks for mathematics.

For science:

- The 1999 (revised 2004) KS3 science curriculum with the TIMSS-2007 Assessment Framework
- The 2007 KS3 curriculum with the TIMSS-2011 Assessment Framework
- The 1999 (revised 2004) KS4 science curriculum with the PISA 2006 Assessment Framework.

The item rating exercise used a subset of PISA and TIMSS items drawn from:

- Released PISA mathematics items from 2003 (the most recent previous cycle in which mathematics was a major domain)
- Secure PISA 2012 mathematics items (the current cycle in which mathematics is again the major domain, with new items as well as trend items)
- Released PISA 2006 science items (the latest cycle in which science was a major domain)
- Released TIMSS-2007 mathematics and science items (at the time of the study, the most recently published cycle of TIMSS)
- Secure TIMSS-2011 mathematics and science items (the current cycle, reported in December 2012).

These items were, generally, newer than those used in the original study. In addition, the original study also used only released items (i.e. those already in the public domain). The current study used a mixture of released and secure items, giving a larger pool (65 for mathematics and 60 for science) from which items were selected to be representative of the

full set of PISA and TIMSS items. The researchers then devised rating instruments based on those used in the original study.

Curriculum comparisons

The areas of assessment were extracted from the published framework for each international study. These were then used to generate spreadsheets listing the areas of assessment against the content of the national curriculum in England at the time of each international study. Where there was no direct match, raters were asked to consider if the area was taught in a previous key stage or in a different subject, or was not in the national curriculum at all.

The raters completed the matching exercise individually but discussed the results at a resolution meeting. Discussion was particularly focussed on areas where there was some variation in interpretation.

Assessment item ratings

The raters were asked to make judgements about seven aspects of each of the mathematics and science items. Overall, they were asked to consider whether the item would be generally appropriate for their students. They were also asked to rate each item according to the percentage of students they estimated would be familiar with the concept, context, format and item type. The final two rating criteria were the subject demand and reading demand. In each case the rating indicated the percentage of students for whom the demand of the item would be suitable.

The raters initially evaluated the items individually but discussed the results at a resolution meeting. The raters expressed the reasons for any individual ratings that varied and then amended their ratings, as appropriate, in light of the discussion that developed.

Results

Comparisons of TIMSS assessment frameworks with the national curriculum

Mathematics

The raters agreed that the content of the TIMSS assessment framework was covered by the relevant national curriculum.

Some of the topics mentioned in the TIMSS framework were not listed explicitly in the national curriculum, but were sufficiently implied for them to be considered as covered. The national curriculum did not re-state basic elements of the curriculum from earlier key stages but they are reflected in more advanced aspects of the topic.

Science

Table 1 indicates that the majority of the TIMSS-2007 framework was covered by the relevant national curriculum in England. Some areas were covered (partially or wholly) in an earlier key stage or in geography (rather than science lessons), so would have been studied by students at some point in their schooling to date.

The TIMSS framework gave specific exemplars of the application of scientific concepts in the real world which may not be familiar or relevant to students in England. For example, while students may study the TIMSS-2007 assessment area of 'Earth's resources, their use and conservation', they would be unlikely to have studied the exemplar of desalination which is more applicable to students living in countries with low rain fall.

Table 1 Science curriculum match for TIMSS-2007 and national curriculum 1999 (revised 2004)

TIMSS-2007 (85 assessment areas in total)	Number of assessment areas
Studied by the relevant age	79
Not in England's KS3 national curriculum at all	6
Partially covered in the science KS3 national curriculum (some parts unfamiliar to students in England)	10
Partially covered by the science KS3 national curriculum and partially covered in previous key stage	1
Covered in previous key stage only	1
Covered in the KS3 geography national curriculum only	5
Partially covered by the KS3 geography national curriculum (some parts unfamiliar to students in England)	1

The scientific competencies and areas of knowledge assessed by TIMSS 2011 were the same as those listed for TIMSS 2007 above. However, slightly fewer (69) specific areas for assessment were identified in the TIMSS 2011 framework.

Table 2 shows that the majority of the TIMSS 2011 framework was covered by the relevant national curriculum in England.

Table 2 Science curriculum match for TIMSS-2011 assessment areas covered by the KS3 national curriculum in England

TIMSS-2011 (69 assessment areas in total)	Number of assessment areas
Studied by the relevant age	68
Not in KS3 national curriculum at all	1
Partially covered in the science KS3 national curriculum (some parts unfamiliar to students in	0

England)	
Partially covered by the science KS3 national curriculum and partially covered in previous key stage	1
Covered in previous key stage only	0
Covered in the KS3 geography national curriculum only	2
Partially covered by the KS3 geography national curriculum (some parts unfamiliar to students in England)	0

Comparisons of PISA assessment frameworks with the national curriculum

It must be noted that PISA assesses students at 15 years of age. Therefore many students are only halfway through their KS4 course and so they may not have yet studied all the topics listed in the KS4 curriculum.

Mathematics

Unlike the TIMSS assessment framework, or a curriculum, the PISA assessment framework did not list mathematical topics. Because its intention was to measure mathematical literacy, it was broken down into content areas, described as 'overarching ideas'.

In order for raters to match the PISA 2003 framework to the national curriculum, the descriptions and explanations of these overarching ideas and cognitive process were used to break these areas down into a further 70 areas.

The 1999 national curriculum to which the framework was matched was a very detailed curriculum and contrasts starkly with the PISA 2003 framework which was much less precise about the content it covered. This made the task of matching the two difficult.

The first rater explained that while the national curriculum programme of study was very detailed:

> *There is no mathematical content mentioned in PISA which is not in [the national curriculum] ... but the scope for PISA to ask KS4 candidates about mathematics which they have not met whilst remaining within its assessment framework is virtually unlimited.*

Despite these difficulties in making the comparisons, all areas of the framework and curriculum were deemed to match.

The 2012 PISA assessment framework had changed since 2003 containing more detail of the content to be assessed. The national curriculum had also changed in style and content since the previous revision in 2004, and had a much less detailed description of content.

Science

Table 3 shows that nearly the entire PISA 2006 framework was covered by the national curriculum in England, but that some areas (from Earth and Space Systems) were studied in geography rather than science lessons.

The TIMSS frameworks (2007 and 2011) are more specific than the PISA framework in detailing the skills and knowledge that can be assessed. The lack of detail in the PISA framework made it more likely to match the national curriculum in England; if the same degree of detail were to be given in the PISA framework there may be more assessment areas which would be rated as only partially covered. It is matched to about the same degree as the TIMSS 2011 framework.

Table 3 Science curriculum match PISA 2006 and national curriculum 1999 (revised 2004)

PISA 2006 (40 assessment areas in total)	Number of assessment areas
Studied by the relevant age	40
Not in KS4 national curriculum at all	0
Partially covered in the science KS4 national curriculum (some parts unfamiliar to England's students)	0
Partially covered by the science KS4 national curriculum and partially covered in previous key stage	0
Covered in previous key stage only	0
Covered in the KS4 geography national curriculum only	4
Partially covered by the KS4 geography national curriculum (some parts unfamiliar to England's students)	1

.

Item analysis

Table 4 below shows the overall ratings for the TIMSS and PISA mathematics items across all nine rating categories. Table 5 shows the overall ratings for the TIMSS and PISA science items across all seven rating categories. Tables 6 to 11 show the ratings broken down by domain

Table 4 Overall ratings of mathematics items' familiarity and suitability

			Familiarity of				Suitability of		Calculator use	
		General appropriateness	Concept	Context	Format	Item type	Subject demand	Reading demand	Calculator use appropriate	Familiarity of having calculator for such items
TIMSS (65 items)	Mean rating	4.20	4.25	4.12	4.38	4.54	4.08	4.43	3.78	4.88
	Standard deviation	0.94	0.93	1.09	0.81	0.69	0.92	0.76	1.04	0.58
PISA (65 items)	Mean rating	4.24	4.47	4.19	4.46	4.73	4.39	4.15	3.91	4.92
	Standard deviation	0.95	0.72	0.85	0.76	0.57	0.81	0.87	1.06	0.46
TIMSS (65 items)	Percentage rated 4 or 5	81.5%	85.4%	73.1%	82.3%	90.0%	77.7%	88.5%	47.7%	96.2%
	Percentage rated 3 to 5	94.6%	93.1%	91.5%	98.5%	99.2%	93.1%	97.7%	96.9%	96.2%
PISA (65 items)	Percentage rated 4 or 5	80.0%	88.5%	77.7%	88.5%	95.4%	82.3%	78.5%	52.3%	96.9%
	Percentage rated 3 to 5	95.4%	99.2%	96.9%	97.7%	99.2%	98.5%	95.4%	96.9%	99.2%

Table 5 Overall ratings of science items' familiarity and suitability

		General appropriateness	Familiarity of				Suitability of	
			Concept	Context	Format	Item type	Subject demand	Reading demand
TIMSS (60 items)	Mean rating	3.00	3.37	3.28	4.02	4.16	2.90	3.60
	Standard deviation	1.27	1.46	1.39	1.15	1.04	1.27	1.11
PISA (60 items)	Mean rating	2.89	2.85	3.02	3.62	4.10	2.99	2.79
	Standard deviation	1.25	1.44	1.28	1.04	1.09	1.31	1.25
TIMSS (60 items)	Percentage rated 4 or 5	35.0%	49.2%	45.8%	72.5%	75.8%	31.7%	57.5%
	Percentage rated 3 to 5	62.5%	79.2%	78.3%	90.8%	92.5%	64.2%	84.2%
PISA (60 items)	Percentage rated 4 or 5	30.3%	37.7%	36.1%	60.7%	81.1%	35.2%	28.7%
	Percentage rated 3 to 5	61.5%	59.8%	65.6%	89.3%	92.6%	64.8%	59.0%

Average ratings of overall appropriateness

Raters were asked "Would this item be generally appropriate for your students?" at an equivalent age with a rating of 1 'not at all appropriate' and 5 'very appropriate'. A rating of 3 means appropriate to half the students, 4 means appropriate for more than half/about 75 per cent of students, 5 all students.

Mathematics

The items were rated very similarly for the two international studies, with a mean rating of 4.2 out of 5, Eighty per cent of items were given a rating of 4 or 5, and almost all (95 per cent) were rated 3 to 5.

Science

The overall mean ratings for the general appropriateness of the TIMSS and PISA items are similar (3.0 and 2.9 out of 5 respectively). For both international studies around 62 per cent of items were given a rating of 3 to 5.

Average ratings for familiarity

Mathematics

Overall a high level of familiarity was reported for the items in both TIMSS and PISA. Table 4 shows that the mean ratings were all above 4 out of 5 for all aspects of familiarity.

Science

The raters felt that 93 per cent of items in both the TIMSS and PISA surveys had an item type that would be familiar to more than half of students (ratings of 3 to 5 with averages of 4.1 overall). Familiarity of format was also rated favourably, although slightly less so for PISA (TIMSS mean rating: 4.0; PISA mean rating: 3.6). The lower PISA value is due to the presence of lengthy introductory text which was deemed sometimes partially or completely irrelevant to the test item(s).

The concept and context of the items in the international studies were rated as less familiar to students in England than item type and format. The PISA ratings (concept: 2.9; context: 3.0) were again a little less favourable than those for TIMSS (concept: 3.4; context: 3.3). Again this is attributed to the quantity or complexity of text presented as a stimulus in PISA needed to understand the background of the question. Nonetheless, it should be noted that even the lowest rating (familiarity of concept for PISA items) was close to the midpoint of the scale and denotes that 60 per cent of items were familiar to more than half of students (ratings of 3 to 5).

Average ratings for subject demand

Mathematics

Of the PISA items, 82 per cent were rated as having a mathematical demand that about 75 per cent of students upwards would be familiar with (ratings of 4 or 5). For the TIMSS items the percentage was slightly lower at 78 per cent.

Science

The subject demand ratings for TIMSS and PISA were similar (2.9 and 3.0 respectively). Sixty-four percent of TIMSS items and 65 per cent of PISA items were thought to be appropriate for more than half of students (ratings of 3 upwards).

Average ratings for reading demand

Mathematics

For mathematics, TIMSS and PISA items were both rated as having a mean reading demand suitable for at least three quarters of students. Of the TIMSS items, 89 per cent were rated as having a reading demand suitable for at least three quarters of students (ratings of 4 or 5), compared to 79 per cent of PISA items.

Science

Eighty-four per cent of the TIMSS items were rated as having a suitable reading demand for half or more of students (ratings of 3 to 5, mean rating 3.6). However, the equivalent proportion of PISA items was 59 per cent and the mean rating (2.8) was the lowest of all the overall ratings. About half of the comments in the rating spreadsheets referred to the length and/or the perceived irrelevancy of the introductory text.

Item ratings by curriculum domain

Table 6 Mathematics item ratings by TIMSS curriculum domains

		General appropriateness	Familiarity of				Suitability of		Calculator use	
			Concept	Context	Format	Item type	Subject demand	Reading demand	Calculator use appropriate	Familiarity of having calculator for such items
Number (19 items)	Mean rating	4.13	4.47	3.89	4.39	4.55	4.32	4.50	3.82	4.61
	Standard deviation	1.12	0.89	1.35	0.86	0.72	0.90	0.73	1.33	1.03
Algebra (22 items)	Mean rating	4.34	3.95	4.36	4.48	4.59	3.84	4.48	3.68	5.00
	Standard deviation	0.71	0.86	0.75	0.73	0.66	0.83	0.66	0.86	0.00
Geometry (13 items)	Mean rating	4.12	4.27	4.04	4.19	4.46	4.00	4.35	3.96	5.00
	Standard deviation	0.95	0.87	1.18	0.90	0.71	0.85	0.94	0.96	0.00
Data and Chance (11 items)	Mean rating	4.14	4.45	4.14	4.36	4.50	4.27	4.32	3.68	5.00
	Standard deviation	1.04	1.10	1.04	0.79	0.74	1.12	0.78	0.95	0.00

Table 7 Mathematics item ratings by PISA content categories

		General appropriateness	Familiarity of				Suitability of		Calculator use	
			Concept	Context	Format	Item type	Subject demand	Reading demand	Calculator use appropriate	Familiarity of having calculator for such items
Change and Relationships 15 items	Mean rating	4.17	4.10	3.83	4.40	4.60	4.00	3.90	3.70	4.93
	Standard deviation	0.87	0.84	0.95	0.77	0.67	0.95	0.84	1.12	0.37
Space and Shape 14 items	Mean rating	4.29	4.36	4.21	4.50	4.79	4.29	4.79	4.04	4.93
	Standard deviation	0.76	0.73	0.83	0.64	0.50	0.81	0.42	0.96	0.38
Quantity 14 items	Mean rating	4.11	4.89	4.61	4.64	4.79	4.89	4.29	4.18	4.86
	Standard deviation	1.34	0.31	0.63	0.73	0.63	0.31	0.76	1.12	0.76
Uncertainty and Data 18 items	Mean rating	4.28	4.58	4.17	4.33	4.72	4.44	3.69	3.81	4.94
	Standard deviation	0.85	0.65	0.88	0.86	0.51	0.81	0.92	0.95	0.33

Table 8 Mathematics item ratings by TIMSS cognitive domains

		General approp-riateness	Familiarity of				Suitability of		Calculator use	
			Concept	Context	Format	Item type	Subject demand	Reading demand	Calculator use appropriate	Familiarity of having calculator for such items
Knowing (20 items)	Mean rating	4.53	4.43	4.58	4.70	4.75	4.40	4.65	3.53	4.93
	Standard deviation	0.64	0.87	0.68	0.56	0.54	0.87	0.58	0.85	0.47
Applying (27 items)	Mean rating	3.96	4.17	3.87	4.28	4.41	3.94	4.39	3.83	4.78
	Standard deviation	1.01	1.08	1.20	0.83	0.74	0.98	0.79	1.22	0.79
Reasoning (18 items)	Mean rating	4.19	4.19	4.00	4.17	4.50	3.94	4.25	3.97	5.00
	Standard deviation	1.04	0.75	1.17	0.91	0.74	0.83	0.84	0.91	0.00

Table 9 Mathematics item ratings by PISA process categories for PISA 2012

		General approp-riateness	Familiarity of				Suitability of		Calculator use	
			Concept	Context	Format	Item type	Subject demand	Reading demand	Calculator use appropriate	Familiarity of having calculator for such items
Formulate (12 items)	Mean rating	4.21	4.13	3.88	4.25	4.67	4.13	4.25	4.42	4.83
	Standard deviation	0.98	0.80	0.99	0.94	0.56	0.95	0.99	0.88	0.56
Employ (15 items)	Mean rating	4.30	4.60	4.30	4.40	4.67	4.57	3.93	4.37	5.00
	Standard deviation	0.70	0.72	0.79	0.77	0.66	0.77	0.87	0.89	0.00
Interpret (11 items)	Mean rating	4.32	4.91	4.45	4.64	4.86	4.86	3.82	3.41	4.82
	Standard deviation	0.89	0.29	0.74	0.49	0.35	0.35	0.80	0.80	0.85

Table 10 Science item ratings by TIMSS curriculum domains

		General appropriateness	Familiarity of				Suitability of	
			Concept	Context	Format	Item type	Subject demand	Reading demand
Biology (21 items)	Mean rating	2.81	3.21	3.19	4.10	4.24	2.81	3.38
	Standard deviation	1.194	1.32	1.33	0.88	0.91	1.25	1.06
Chemistry (12 items)	Mean rating	3.29	3.83	3.63	4.33	4.42	3.08	3.96
	Standard deviation	1.52	1.63	1.56	1.09	1.06	1.59	1.00
Earth Science (12 items)	Mean rating	2.88	2.67	2.71	3.38	3.63	2.54	3.46
	Standard deviation	1.12	1.63	1.52	1.53	1.21	1.25	1.28
Physics (15 items)	Mean rating	3.13	3.77	3.60	4.17	4.27	3.17	3.73
	Standard deviation	1.28	1.14	1.07	1.05	0.94	0.95	1.08

Table 11 Science item ratings by TIMSS cognitive domains

		General appropriateness	Familiarity of				Suitability of	
			Concept	Context	Format	Item type	Subject demand	Reading demand
Knowing (20 items)	Mean rating	3.15	3.38	3.35	4.65	4.80	3.18	4.18
	Standard deviation	1.35	1.64	1.53	0.62	0.46	1.43	0.78
Applying (20 items)	Mean rating	3.05	3.36	3.34	3.80	3.91	2.77	3.52
	Standard deviation	1.311	1.46	1.43	1.27	1.10	1.14	1.19
Reasoning (20 items)	Mean rating	2.78	3.36	3.14	3.58	3.75	2.75	3.06
	Standard deviation	1.12	1.27	1.17	1.18	1.11	1.20	1.04

Ratings by curriculum domain for overall appropriateness familiarity and subject demand

Mathematics

There was little difference in the ratings of general appropriateness and all showed that TIMSS and PISA items were appropriate.

Science

The ratings for the general appropriateness by TIMSS curriculum domain lie around the midpoint with Earth Science items being rated as appropriate for a little less than half of students.

The familiarity ratings for Earth Sciences were lower than those for the other three curriculum domains in all four familiarity categories possibly because the content is covered in geography lessons

Discussion

Curriculum match

The national curriculum for mathematics was found to assess the same content and skills as laid out in the mathematics assessment frameworks of both the PISA and TIMSS surveys.

Similarly the national curriculum for science was found to broadly assess the same content and skills as laid out in the assessment frameworks of both the PISA and TIMSS surveys. It is possible that the small differences in content could make the international science assessments appear harder than the national science assessments with TIMSS-2007 perhaps the most difficult but there is little evidence for this, and England continues to do relatively well in the surveys in science, so this is unlikely to be an issue.

Assessment item ratings

Almost all TIMSS and PISA mathematics items were deemed familiar and suitably demanding for at least half of students.

The reading demand of both studies was thought to be broadly appropriate but the reading demand of PISA science items was thought to be less suitable than the reading demand of the TIMSS science items. The PISA items are more likely to make use of a lengthy stimulus material. It must be borne in mind that PISA aims to assess students' readiness for the adult word and so aims to use contexts such as newspaper articles etc. which will have higher reading demand. Care will need to be taken if reading demand is reduced so that that it does not change what PISA is assessing.

There was little difference between the two international studies with regard to subject demand which was considered appropriate. The science items were also considered generally familiar, although the format, context and concept of the TIMSS items were judged to be more familiar than that of the PISA items. The amount of text in the PISA items was mostly responsible for this difference.

Conclusions and implications

The comparison of the assessment frameworks and test items reveals that there is a high degree of match both between TIMSS and PISA and with the national curriculum in England. Furthermore the level expected (subject demand), the familiarity of format, contexts and concepts were judged to be similar and appropriate for English learners at the relevant stages.

Reading demand was judged to be higher, especially in PISA. This could introduce some invalid, non-domain relevant distortion into the international survey results but it is likely that learners in other countries also experience this additional demand due to the need for the international surveys to provide the contexts and required data in the questions.

It is also clear that the reading demand appears to be equally high in the mathematics and the science items, yet England performs relatively better in the PISA science items than the PISA mathematics items. This would suggest that the lower performance in PISA mathematics, or differential performance between the two studies is unlikely to be caused by the reading demands.

It is noted that both TIMSS and PISA are making efforts to reduce this demand and continually improve and refine their assessments but the international surveys need to continue to monitor and reduce the amount of non-domain demand, especially in PISA science.

References list

Department for Children, Schools and Families and Qualifications and Curriculum Authority (2007a). 'Mathematics: programme of study for key stage 3 and attainment targets.' In: *The National Curriculum: Statutory Requirements for Key Stages 3 and 4*. London: QCA and DCSF [online]. Available: http://media.education.gov.uk/assets/files/pdf/q/mathematics%202007%20programme%20of%20study%20for%20key%20stage%203.pdf [3 May, 2013].

Department for Children, Schools and Families and Qualifications and Curriculum Authority (2007b). 'Mathematics: programme of study for key stage 4 and attainment targets.' In: *The National Curriculum: Statutory Requirements for Key Stages 3 and 4*. London: QCA and DCSF [online]. Available:http://media.education.gov.uk/assets/files/pdf/q/mathematics%202007%20programme%20of%20study%20for%20key%20stage%204.pdf [3 May 2013]

Department for Children, Schools and Families and Qualifications and Curriculum Authority (2007c). 'Science: programme of study for key stage 3 and attainment targets.' In: *The National Curriculum: Statutory Requirements for Key Stages 3 and 4*. London: QCA and DCSF [online]. Available: http://media.education.gov.uk/assets/files/pdf/q/science%202007%20programme%20of%20study%20for%20key%20stage%203.pdf [3 May, 2013].

Department for Education and Skills (2004). *Programme for International Student Assessment (PISA) 2003: England Sample And Data* (SFR 47/2004). London: DfES.

Department for Education and Skills and Qualifications and Curriculum Authority (2004a). 'Key stage 3 programmes of study: mathematics.' In: *The National Curriculum: Handbook for Secondary Teachers in England. Key Stages 3 and 4*. London: QCA and DfES [online]. Available: https://www.education.gov.uk/publications/eOrderingDownload/QCA-04-1374.pdf [3 May, 2013].

Department for Education and Skills and Qualifications and Curriculum Authority (2004b). 'Key stage 4 programmes of study: mathematics.' In: *The National Curriculum: Handbook for Secondary Teachers in England. Key Stages 3 and 4.* London: QCA and DfES [online]. Available: https://www.education.gov.uk/publications/eOrderingDownload/QCA-04-1374.pdf [3 May, 2013].

Department for Education and Skills and Qualifications and Curriculum Authority (2004c). 'Key stage 3 programmes of study: science.' In: *The National Curriculum: Handbook for Secondary Teachers in England. Key Stages 3 and 4.* London: QCA and DfES [online]. Available: https://www.education.gov.uk/publications/eOrderingDownload/QCA-04-1374.pdf [3 May, 2013].

Department for Education and Skills and Qualifications and Curriculum Authority (2004d). 'Key stage 4 programmes of study: science.' In: *The National Curriculum: Handbook for Secondary Teachers in England. Key Stages 3 and 4.* London: QCA and DfES [online]. Available: https://www.education.gov.uk/publications/eOrderingDownload/QCA-04-1374.pdf [3 May, 2013].

Mullis, I.V.S., Martin, M.O., Ruddock, G.J., O'Sullivan, C.Y., Arora, A. and Erberber, E. (2005). *TIMSS 2007 Assessment Frameworks* Chestnut Hill, MA: Boston College, International Study Center, Lynch School of Education [online]. Available: http://timss.bc.edu/timss2007/PDF/T07_AF.pdf [3 May, 2013].

Mullis, I.V.S., Martin, M.O., Ruddock, G.J., O'Sullivan, C.Y. and Preuschoff, C. (2009). *TIMSS 2011 Assessment Frameworks*. Chestnut Hill, MA: Boston College, International Study Center, Lynch School of Education [online]. Available: http://timssandpirls.bc.edu/timss2011/downloads/TIMSS2011_Frameworks.pdf [3 May, 2013].

OECD (2004). *The PISA 2003 Assessment Framework: Mathematics, Reading, Science and Problem Solving Knowledge and Skills.* Paris: OECD Publishing [online]. Available: http://www.oecd.org/edu/preschoolandschool/programmeforinternationalstudentassessmentpisa/33694881.pdf [3 May, 2013].
OECD (2006). *Assessing Scientific, Reading and Mathematical Literacy: a Framework for PISA 2006.* Paris: OECD Publishing [online]. Available: http://www.oecd.org/pisa/pisaproducts/pisa2006/37464175.pdf [3 May, 2013].

OECD (2010). *PISA 2012 Mathematics Framework* (draft). Paris: OECD Publishing [online]. Available: http://www.oecd.org/pisa/pisaproducts/46961598.pdf [3 May, 2013].

Ruddock, G., Clausen-May, T., Purple, C. and 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 [online]. Available: http://webarchive.nationalarchives.gov.uk/20130401151715/https://www.education.gov.uk/publications/eOrderingDownload/RR772.pdf [3 May, 2013].

Sturman, L., Burge, B., Cook, R. and Weaving, H. (2012). *TIMSS 2011: Mathematics and Science Achievement in England.* Slough: NFER [online]. Available: http://www.nfer.ac.uk/publications/TMEZ01 [3 May, 2013].